PARTIAL ADVANCE INFORMATION IN A CHOICE REACTION TASK

BY J. ALFRED LEONARD

Medical Research Council's Applied Psychology Research Unit, Cambridge

An experiment was carried out to study the effect of partial advance information on performance. Reaction time and accuracy measures were compared for three conditions: a six-choice task, a three-choice task, and a set of experimental conditions in which a two-choice task preceded two three-choice tasks by various durations of time. Tests under all conditions were carried out after varying amounts of practice. It was found that at all stages of practice a difference between performance on the six-choice task and on the three-choice task was maintained. As the separation between the two-choice and three-choice task was varied from 0·04 sec. to 0·50 sec. in the experimental conditions, performance became increasingly similar to, and finally identical with, that under the control three-choice condition. The relationship between duration of forewarning and performance was, however, not a simple one. The results are discussed with reference to traditional and current views on information processing.

I. INTRODUCTION

The experiment to be described stemmed originally from work on anticipation in sensory-motor skills (Leonard, 1953, 1954), and was given an additional impetus by Hick's (1952) paper on the rate of gain of information. In discussing the relationship between 'receptor-effector' anticipation and anticipation based on past experience, I stated that the latter was frequently not entirely adequate to prepare responses, because 'very often there is a residue of uncertainty, a moment to moment randomness in the environment'. Receptor-effector anticipation enables one to reduce this residue of uncertainty just before a point of necessary action is reached, so that performance does not merely become faster but also smoother, responses merging one into the other without hesitation or raggedness. It was possible to show this in two experimental situations in which performance was compared when a new signal was given either at, or before, a choice point was reached in serial choice reaction tasks. This signal was thought of as providing the final item of information necessary to complete a given response, the rest of the information being provided by instructions and the subject's eventual experience of the task; in other words, anticipation based on past experience. From those experiments it seemed only a natural step to study the effects of providing the final item not in one but in several stages, to study the effects of partial advance information. The influence of Hick's work arises from the fact that in order to account for his finding of a constant rate of gain of information, he suggested as a reasonable possibility an internal process of progressive binary classification of information (1952, p. 22). It will be seen that the experimental situation chosen for the study of partial advance information was essentially one in which information could be presented to the subject in a progressively classified manner.

A method of doing this in a serial task comparable to those used in the earlier experiments would have been desirable from the point of view of continuity of research. But since practically nothing was known about the effects of partial advance information in skilled performance it seemed wiser to study these first in a setting free from possibly masking serial effects. As my experimental situation I took, therefore, a straightforward

six-choice task. It is possible to think of this as a two-choice task followed by a three-choice task. Essentially the experiment was a comparison of three conditions: a six-choice task, a three-choice task, and a set of conditions in which a two-choice preceded a three-choice task by varying amounts of time.

Now the last set of conditions can be thought of as one in which information is presented in a progressively classified manner. It was just such a system of *internal* progressive classification which Hick suggested as a possible mechanism to account for his finding of a constant rate of gain of information. There were two main questions to be answered: first, how far does the two-choice signal have to precede the three-choice signal for a subject to be able to respond to the latter as efficiently as in a pure three-choice situation? And secondly, what is the relationship between performance when all the information necessary is presented simultaneously and when it is presented successively, i.e. between the pure six-choice task, and the progressively classified task?

II. Method and procedure

Subject. The data to be presented were those obtained by myself acting as the subject.

Apparatus. The display consisted of six small neon bulbs stuck through a blackened piece of metal, so as to form an inverted V, with three bulbs to each arm. The display panel was shielded to reduce glare and was placed about 3·5 ft. from *S*, far enough to allow all bulbs to be seen easily when looking at the centre of the V. The control consisted of six micro-switches with paxolin flaps, raised at an angle of about 50°. The hands rested on an inclined panel so that the keys had to be pressed in the direction of the palm, i.e. a fairly natural clenching movement. Circuits were so constructed that the apparatus could deliver sequences of 48 stimuli in any order desired.

Conditions. The unit of presentation was a trial in which one had to respond to 48 signals. Each new signal appeared 1·5 sec. after a correct response had been made. There were two *control* conditions, namely six-choice and three-choice; and a group of *experimental* conditions in which a two-choice preceded a three-choice by varying amounts of time.

In a trial under the *six-choice condition* all the six lamps would light up together. After 0·1 sec. one of the six lights would go out.

In a trial under the *three-choice condition* the three lamps of only *one* set came on together, this set, i.e. left or right, being the same one throughout the trial. After an interval of 0·1 sec. one of the three lights went out.

In a trial under the *experimental condition* the lights of only one set came on together at any one time, but it could be either of the two sets. After an interval which was constant for any given trial, but which could be varied between trials, one light went out.

The coming on of the lights will be referred to as the *warning signal*, going out of one light as the *specific signal*, the time between as the duration of *forewarning*.

It will be seen that under the control conditions the warning signal indicated that a specific signal would have to be responded to in 0·1 sec. In the experimental condition the warning signal indicated the one of the two sets of three lights in which a specific signal would occur after a known interval. It was by this means that information was presented in two stages.

Ideally it would have been desirable to include trials in which the control conditions were given under all durations of forewarning. This would have added considerably to the times for both the running of the experiment and analysis of data. The 0·1 sec. interval for the two control conditions was chosen as a standard after extensive preliminary trials had shown that varying the duration of the forewarning in the control conditions had no facilitating effect. If anything, durations beyond 0·3 sec. tended to retard performance. (See Klemmer (1956) for recent evidence.)

Under all conditions the *correct response* was to press the key spatially corresponding to the light which had gone out. This resulted in putting out the remaining lights of the display. After an interval of 1·5 sec. the next warning signal came on. *Pressing a wrong key* caused a buzzer to sound; there was no effect on the visual display. All the keys could be pressed simultaneously so that if a correct and a wrong key were pressed together, the light went out and the buzzer sounded.

Recording. This was carried out with N. T. Welford's SETAR (1952). Time readings were taken to the nearest 0·01 sec. In all conditions, except that of the experimental condition with the shortest forewarning, the intervals recorded were as follows: (1) between onset of warning signal and onset of specific signal (forewarning); (2) between onset of specific signal and subject's first response (reaction time); (3) between subject's response and onset of warning signal (rest interval).

Sequences. The same basic sequence of 48 signals constituting one trial was used throughout the major part of the experiment. This sequence had been specially constructed with the following properties: (1) there were an equal number of signals for each category of choice, i.e. in any trial each light would appear an equal number of times; (2) runs of more than two identical signals did not appear (this applied to the sequence when used for the six-choice and the experimental conditions only); (3) diagram frequencies were balanced as far as possible. The basic sequence is reproduced below:

125663445326433614315514654242116636213532524152

Six six-way single pole switches enabled the sequence to be permuted when required, so that the subject need not have the same succession of signals twice, while the basic properties of the sequence were maintained. For the three-choice condition the sequence was transformed by always putting two of the permuting switches on the same position. In the earlier stages of the experiment a specially designed three-choice sequence had been used but the later method had been shown to produce the same results, and saved a great deal of time.

Circuits. The main components were electro-magnetic relays of various kinds. The forewarning and rest-interval periods were condenser controlled. The motivating device (see below) had an electronic timing circuit. All the apparatus—except, on some tests, the counter attached to the motivating device—was outside the experimental room.

Scoring. Two scores were obtained: duration and accuracy of each response. The time score taken was the interval between the onset of the specific signal and the subject's first response to it, regardless of whether this was a correct or an error response. It should be noted that since all the main data to be presented are averages obtained from all six fingers under all conditions finger differences are taken care of. The same applies to possible variations among keys and relays. Times were averaged for sets of trials, a trial consisting of 48 responses. Error scores were obtained directly from the SETAR record. Rates of gain were calculated in the manner described by Hick (1952, pp. 25–6). It should be noted that there are two ways of doing this for the experimental conditions: one can think of them either as six-choice tasks with the information given in two stages; or one can think of them as three-choice tasks only, ignoring the warning stage. In the former case one has to consider an input of 2·59 bits and divide the transmitted information by the sum of the forewarning period plus the reaction time observed; in the latter case one treats the hands separately, considers an input of 1·58 bits to each hand, and divides the information transmitted by the reaction time only. Treating the hands separately is perhaps only justified as long as there are no confusions between hands; since this obtained and the reaction times observed for the longer durations of forewarning were closer to the three-choice than to the six-choice times, the rates of gain for the experimental conditions with all but the shortest duration of forewarning were so calculated.

Motivation. Knowledge of results could be provided by means of an all or none indicator. The subject could activate an electromagnetic counter provided he responded correctly below a given but adjustable time limit. The counter could be either in the experimental room or outside with the rest of the apparatus.

Procedure. The over-all plan was to carry out tests at various stages during extended practice. A total of at least 63,000 responses under the various conditions was made during the 8 months of experimentation. The procedure during practice and between test series was not altogether uniform, so that one is perhaps not quite justified in presenting the data in the form of 'learning curves'. The main differences among the procedures were in the per cent of errors tolerated, the order in which the conditions were given, and the different uses of the motivating device. Of these the first is by far the most important and should be clearly stated: for the *first* test (June, after about 9600 responses) I tried to make as few errors as possible but did not reject any trials with high error incidence. In the *second* test (September, after about 26,400 responses) I tried to keep the number of errors constant for all conditions and rejected trials having more than three errors. In the *third* test (November, after about 52,800 responses) I paid no attention to the number of errors made at all. This last test had the most responses and was the most rigorously conducted of the three. Conditions were given in ascending and descending series, i.e. from six-choice through the various experimental conditions to three-choice where left and right hand were used alternately on succeeding trials. Trials were only rejected on the grounds of apparatus failure: four in 168. There was a 3 min. interval between trials, and 15 min. between each series of seven trials. On the

first two days of the third test two ascending and two descending series were given, and on the remaining five experimental days four such series each, a total of 24 complete series, 12 ascending and 12 descending. The error indicator was not used. The motivating device's counter was in the experimental room and the difficulty so adjusted that about 80–90% of the responses produced a click. It should be noted that the level was adjusted according to the condition used.

III. Results

Distributions of reaction times for three conditions of the September test are presented in Fig. 1. They are fairly typical of all distributions obtained. Since error level was constant and low throughout that test a comparison of these distributions is perhaps more meaningful than those obtained at the other tests. They indicate that the use of the

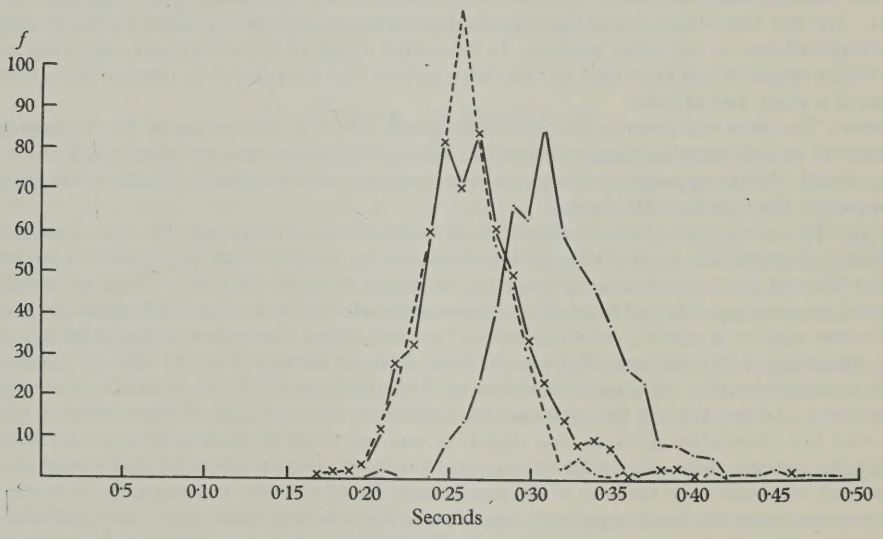


Fig. 1. Reaction time distributions for the September test for the two control conditions and the experimental conditions with 0·5 sec. forewarning. ·——·, six-choice, $N=576$, $m=0{\cdot}319$; ·----·, Exp. 0·50, $N=576$, $m=0{\cdot}262$; ×—— ×, three-choice, $N=576$, $m=0{\cdot}268$.

arithmetic mean of reaction times can be considered as justified. Fig. 1 shows the separation of the distributions for the two control conditions. Table 1 shows the results in terms of numbers of responses per condition, mean reaction time, and number of errors, and information measures, where calculated (see scoring above). The values for the experimental and three-choice conditions are averages for both hands.

Fig. 2 shows reaction time and error percentage curves for the three tests; the horizontal axis is of course an arbitrary scale, with the values for six-choice on the left, followed by those for the experimental conditions, and those for the three-choice on the right-hand side. This seemed a convenient way of presenting the three sets of data graphically; continuity between six-choice and three-choice is at this time an open question. It will be seen that very much the same pattern of results was obtained in all three tests.

(1) There was a clear difference between performance under the two control conditions.

(2) In the experimental condition there was a measurable effect on reaction time when the duration for forewarning was varied in the sense that broadly speaking as this duration

was increased reaction time became shorter: (*a*) the largest single effect of reducing the reaction time was obtained with 0·1 sec. forewarning; (*b*) there was some retardation of the effect with forewarning of 0·2 sec.; and (*c*) as the duration of forewarning reached 0·5 sec. reaction time tended to be equal to or better than the three-choice reaction time.

(3) Concerning the information measures it should be noted that: (*a*) on all three tests the rates of gain obtained for the six-choice control condition were higher than those for

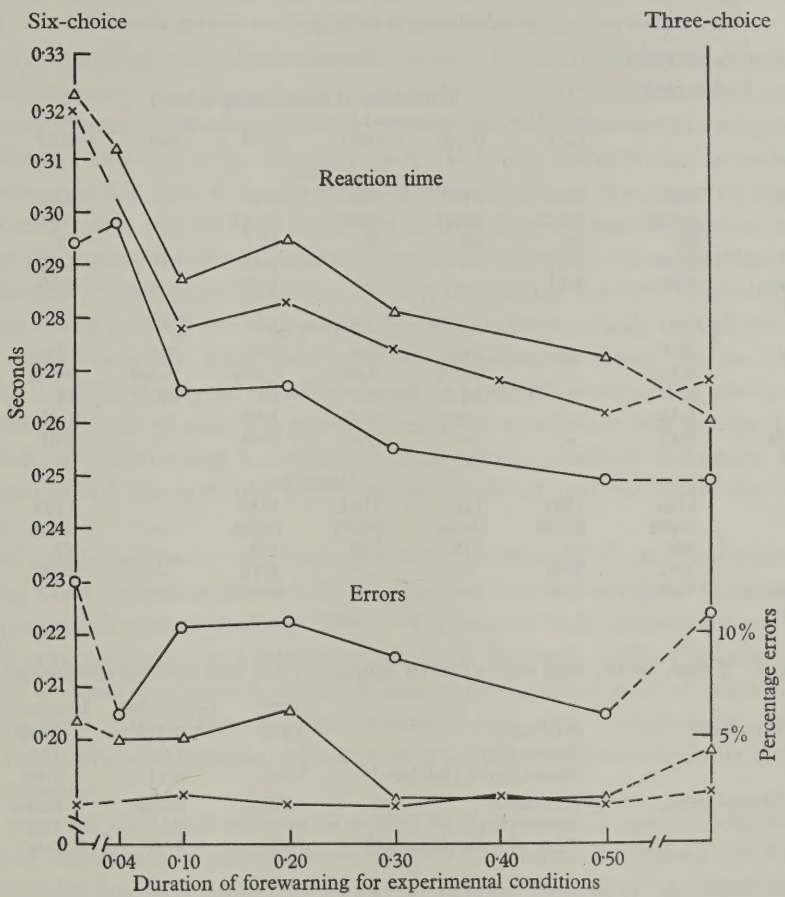


Fig. 2. Curves showing mean reaction times and error percentages for all conditions of the three tests: △——△, June; ×——×, September; ○——○, December.

the three-choice conditions; (*b*) on the last test the rate of gain for the experimental condition with 0·04 sec. forewarning came nearer to that of the six-choice condition; and (*c*) on all three tests rates of gain for the experimental condition with 0·3 sec. forewarning were comparable to those of the three-choice condition, while at 0·5 sec. forewarning rates of gain were higher on all three occasions.

Practice effects. In view of the fact that conditions were not entirely constant throughout the whole course of the experiment, progressive effects among identical conditions on different tests cannot be ascribed solely to practice. Table 2 is a reorganization of some of the data shown in Table 1 to show how reaction times, errors, and rates of gain changed

for the two control conditions on successive tests. Rate of gain, which takes into account both speed and accuracy of performance, is here a useful measure of efficiency. The highest rates of gain were obtained for the September Test. Reaction times were lowest on the December Test, but not low enough to compensate for the higher error level, so that rates of gain were lower than in September.

Table 1. *Results of three tests taken at various times*

(For methods of calculating rates of gain see scoring above)

Condition ...	Control Six-choice	Experimental (Duration of forewarning in sec.)						Control Three-choice
		0·04	0·10	0·20	0·30	0·40	0·50	
					June			
Responses	575	575	576	576	576	—	576	1152
R.T. (sec.)	0·323	0·312	0·287	0·295	0·281	—	0·273	0·261
Errors	35	29	29	36	12	—	12	49
H_T (bits)	2·24	2·27	—	—	1·44	—	1·43	1·29
Rate of gain (bits/sec.)	7·80	7·21	—	—	6·73	—	6·90	6·85
					September			
Responses	576	—	576	528	576	576	576	576
R.T. (sec.)	0·319	—	0·278	0·283	0·274	0·268	0·262	0·268
Errors	11	—	13	9	10	12	11	14
H_T (bits)	2·44	—	—	—	1·46	—	1·47	1·44
Rate of gain (bits/sec.)	8·41	—	—	—	6·96	—	7·31	7·06
					December			
Responses	1104	1102	1152	1104	1152	—	1103	1152
R.T. (sec.)	0·294	0·298	0·266	0·267	0·255	—	0·249	0·249
Errors	140	67	119	116	101	—	67	123
H_T (bits)	1·91	2·26	—	—	1·11	—	1·14	1·02
Rate of gain (bits/sec.)	7·68	7·50	—	—	6·52	—	6·75	6·44

Table 2. *Time, error, and rate of gain scores for the two control conditions only*

		June	September	December
Six-choice	R.T. (sec.)	0·323	0·319	0·294
	Errors (%)	6·1	1·9	12·7
	Rate of gain (bits/sec.)	7·80	8·41	7·68
Three-choice	R.T. (sec.)	0·261	0·268	0·249
	Errors (%)	4·3	2·4	10·7
	Rate of gain (bits/sec.)	6·85	7·06	6·44

IV. Discussion

Before answering the two questions posed at the end of the Introduction it is necessary to draw attention to the major 'incidental' finding: there was, throughout, a clear difference in the results obtained under the two control conditions. The difference in reaction time was rather smaller than that expected (e.g. Woodworth, p. 333, or Hick, p. 15) and, conversely, that in rate of gain was unexpected (Hick, 1952). While this paper was in preparation, a considerable amount of evidence appeared to show that the slopes describing the relationship between entropy and choice time are in turn related to the particular combinations of input and output coding used in a given experiment (for a discussion see Crossman, 1956). The present results show that for a particular input-output coding variation in entropy continued to show an effect even after a considerable amount of practice.

The answer to the first question (namely, how far does the two-choice signal have to precede the three-choice signal for a subject to be able to respond to the latter as efficiently as in a pure three-choice situation?) is best given by reference to the rate of gain score: the best estimate is 0·30 sec. Rather surprisingly, efficiency improved consistently when the separation was further increased up to 0·50 sec. It is interesting to note that May (1917), using the method of controlled association, also had a maximum effect of forewarning when the duration was 0·50 sec. (for a brief description see Woodworth, 1938, pp. 791–3).

The second question concerned the relationship between performance when all the information necessary is presented simultaneously and when it is presented successively. Using the rate of gain score again, it would appear that the measures are comparable only with the shortest duration of forewarning used. (It will be recalled that to make this kind of comparison one has to add the duration of forewarning to the observed reaction time and consider an input of 2·59 bits. Since the observed reductions in reaction time due to increases in the duration of forewarning are small in relation to that duration the rates of gain will become increasingly smaller.) It is worth noting that the shortest duration of forewarning it was possible to achieve with the apparatus available turned out to be close to the observed absolute difference between the reaction times for the two control conditions. It would clearly be of some interest to establish whether the above-mentioned comparability of rates of gain was purely fortuitous or whether it is always found when the duration of forewarning is similar to that of the absolute difference in reaction times, regardless of the absolute values of the observed reaction times for two control conditions.

A second incidental result was the consistent retardation of the facilitating effect of forewarning when its duration was 0·20 sec. Similar results have since been presented by Elithorn (personal communication, 1956). I interpret this retardation as a possible instance of refractoriness (for a discussion see A. T. Welford, 1952). The point of interest here is that whereas in classical refractoriness experiments a pair of stimuli has to be responded to by a pair of responses, in the present case only a single overt response was required. The findings suggest that refractoriness may not be due exclusively to the making of a first response.

Future work on partial advance information in this area will have to take into account the effects of differences in possible input-output relationships. It may not be necessary to go to the same length in training but to concentrate instead on the three durations of forewarning mentioned above.

Grateful acknowledgements are due to Mr A. T. Welford, Director of the Nuffield Research Unit into Problems of Ageing, for advice and facilities generously offered; to Dr N. T. Welford for services psychological and electronic; to Dr W. E. Hick for stimulation and reinforcement. This work was carried out while the author was in receipt of a stipend from the Medical Research Council and the Kenneth Craik Award for 1952–3. It was contained in A.P.U. 217/54.

References

Crossman, E. R. F. W. (1956). The information capacity of the human operator in symbolic and non-symbolic control processes. In *The Application of Information Theory to Human Operator Problems*. London: M.O.S. Report WR(D) 2/56.

Hick, W. E. (1952). On the rate of gain of information. *Quart. J. Exp. Psychol.* **4**, 11–26.

Klemmer, E. T. (1956). Time uncertainty in simple reaction time. *J. Exp. Psychol.* **51**, 179–84.

Leonard, J. A. (1953). Advance information in sensorimotor skills. *Quart. J. Exp. Psychol.* **5**, 141–9.

Leonard, J. A. (1954). An experiment with occasional false information. *Quart. J. Exp. Psychol.* **6**, 79–85.

May, M. A. (1917). The mechanism of controlled association. *Arch. Psychol.* **39**.

Welford, A. T. (1952). The 'psychological refractory period' and the timing of highspeed performance, a review and a theory. *Brit. J. Psychol.* **43**, 2–19.

Welford, N. T. (1952). The electronic digital recording machine, the SETAR. *J. Sci. Instrum.* **29**, 1–4.

Woodworth, R. S. (1938). *Experimental Psychology*. New York: Holt.

(*Manuscript received* 5 *November* 1956)

FACTORS AFFECTING INDUCTIVE PREDICTIONS

By R. T. GREEN

University College, London

Two guessing tasks which appear highly similar in nature nevertheless evoke opposite trends in predictive behaviour. Attempts to generate both sets of results from the same simple principles fail because of the different implicit assumptions the subject makes about the nature of the task. It would seem that one of the assumptions that the subject tends to make when required to employ inductive reasoning, at least in the laboratory situation, is that a certain amount of stimulus variability is to be expected.

The theoretical implications are that, far from being contradictory or mutually exclusive, the gestalt and behaviourist accounts of the learning process are complementary. Only by recognizing that some aspects of the learning process are continuous while other aspects are discrete and that both S–R learning at an unverbalized level and conscious hypothesis testing go on side by side, is it possible to make sense of the more complex type of learning situation where a temporal pattern is involved.

An attempt to analyse pre-solution guessing behaviour tends to favour a non-continuous hypothesis, but in certain respects the question is still open to a more sophisticated interpretation.

I. Introduction

Learning theory has long been sharply divided over such questions as the role of reinforcement, and whether the S–R or the S–S paradigm meets the facts. More recently, we have seen these paths converging as a process of mutual adaptation makes the different approaches complementary instead of contradictory. On the surface the continuity-discontinuity controversy appeared to be an instance of outright warfare between the gestalt hypothesis-testing, insightful account of the learning process, and the behaviourist explanation in terms of gradual learning on a trial-by-trial reinforcement basis. Yet, after ploughing through the literature pro and con, experiment and counter-experiment, one is left with the distinct impression that both parties are right. In particular, consider the work of Lawrence (1949, 1950), who found that his rats learned to attend to the *relevant* cues in a discrimination task as demanded by gestalt theory, yet conformed to behaviourist principles in other respects. Prentice (1949), sympathetic to the gestalt view, found that his human subjects learned to discriminate in a way which baffled both continuous and non-continuous types of interpretation. And the earlier behaviourist approach of Spence (1940) runs into difficulties when more subtle discriminations based on transverse patterning are involved, as Bitterman (1953) points out.

Two main factors manage to obscure a great deal which the two viewpoints have in common. There is, first, a tendency for both to say the same thing in different ways. Hull (1952) ended up by leaning so heavily on concepts like the 'pure stimulus act' and 'fractional antedating goal reaction' that 'insight' became indistinguishable from 'implicit trial and error' in any operational sense. The second factor that has helped to give an unjustified impression of incompatibility is that, whereas the behaviourists have based their position on the statistical trends of groups of subjects, the gestalt theory is more concerned to describe what the individual subject is doing. Nowhere is this more apparent than when we employ human subjects for a concept-formation task using the accepted paradigm of reversal in the pre-solution period. The data of Walk (1952) seem to lend

support to the behaviourist contention; the more pre-solution practice that is given on the reversed discrimination, the more difficult it is to establish the final concept. Yet no one disputes the fact that human subjects test hypotheses while searching for a solution. All the behaviourist can do is to deny the relevance of this fact to the actual learning that occurs. At this juncture, the gestalt theorist can turn the tables very neatly: he can show that the result is quite compatible with his theory and, simultaneously, he can convict the behaviourist of an inherent logical contradiction. Accepting that the subject tries out hypotheses, then, the longer the pre-solution period on the reversed discrimination, the greater the probability that the subject will test what is to *become* the correct hypothesis and abandon it. As long as he remains unaware that the task has been changed he is more likely to avoid the correct hypothesis under the new conditions. Thus, the experimental evidence is in no way at variance with the gestalt position. Now for the logical contradiction. It is an accepted condition of this type of 'crucial' experiment that the original learning on the reversed discrimination must not go beyond the pre-solution period (see McClelland, 1943). Yet according to the behaviourist account in terms of a gradual, trial-by-trial acquisition of 'habit strength' there can be no such thing as a pre-solution period. From trial *1* to trial *999* there is no point at which the problem can be said to be solved or unsolved. Habit strength increases logarithmically to an asymptote. Learning is never perfect nor utterly imperfect. Attempts to establish the pre-solution period by showing that scores are not above chance either demonstrate that the gradual acquisition of habit strength theory is wrong or that the statistics are insensitive to the learning that must have occurred. Spence (1952) may protest that behaviourist learning theory was developed to deal with the behaviour of rats and is therefore not proven or disproven by what human subjects do. But behaviourists have only themselves to blame for such a misunderstanding, since no one has been more eager than they to explain human behaviour in S–R terms on every possible occasion. Whenever they are criticized as 'rat' psychologists, their defence is that they are seeking basic principles using lower organisms in the hope of developing a theory to deal with the more complex behaviour shown by human subjects. Spence (1940) is not averse to citing his own experiments on chimpanzees in this context.

One is driven to conclude that both schools are describing, in their own terms and for the most part accurately, different aspects of the same process. There is little doubt that the subject tests hypotheses until he finds one that pays off, or runs out of hypotheses. He continually abandons and modifies these hypotheses in the light of success and failure. He may learn that a certain type of hypothesis tends to pay off better than any other and then concentrate his search in this area. In behaviourist terms, he is learning to attend (respond) to the relevant stimuli and ignore the irrelevant. It is precisely this attending to the relevant stimuli that the gestalt theorists insist is essential for learning to take place. At the same time they overlook the possibility that, although the subject attends to the relevant stimuli, he may be directed to do so by the wrong hypothesis. In which case, it would be quite reasonable to expect some learning to take place in the S–R reinforcement manner although the problem remains unsolved. Errors for the group may therefore tend to show a steady decline. Again, if a subject hits the correct hypothesis, his errors will cease but group results will smooth out this 'insight'. In any case, we never know what process of implicit trial and error leads up to, and hence makes less dramatic, this sudden individual drop in overt errors. Furthermore, as Hull (1920) showed, it is

possible for a subject to respond 100% correctly and yet be unable to verbalize the concept.

All this preamble has been intended to emphasize the need for a more eclectic approach to the learning process if we are to understand some of the more subtle and complex phenomena that keep intruding themselves into the data and upsetting our highly polished theories. Polish is admirable—on timber that has the quality to take it. The experiments to be described pose knotty problems unless one is willing to adopt an open-minded view about what goes on when people learn; that is, modify their responses in the light of past experience. One is never sure just how far back the experience goes which is helping to determine a particular response: one does know that it extends far beyond the beginning of any laboratory experiment. To demonstrate the importance of this influence the binary guessing behaviour of a group of subjects who have to predict a light sequence is contrasted with that of another group who are faced with a very similar binary sequence using verbal material. Another light sequence reveals something of the way a simple alternation habit is learned.

II. Experimental procedure

Method 1. *Light guessing.* The apparatus consisted simply of two 3·5 V. bulbs operated independently by finger keys and supplied by a small dry cell. The light on the right of the subjects was backed by red paper and referred to as the 'red light', while that on the left was designated 'yellow' in a similar fashion.

The subjects, 121 in all, were students at a technical college, aged 15–35 with a mean age of about 23 years. Six groups were involved, two on each of three series to counterbalance for position or colour preferences. The structures of the three series (shown in Table 1) followed those of a previous memory experiment involving numbers and nonsense syllables. It had been hoped in this way to interpret differential recall probabilities in terms of unexpected structural change, but an unexpected variable proved to be more in need of interpretation.

Table 1. *Structure of the three light-guessing series*

	Serial position											
Groups	1	2	3	4	5	6	7	8	9	10	11	12
I	X	O	X	O	X	O	X	O	X	O	X	O
II	X	X	X	O	O	O	O	O	O	X	X	X
III	X	X	X	O	X	X	X	O	O	O	O	O

X, red or yellow light; O, light of alternative colour

Formal instructions to the subjects, run in groups of ten or twenty, were as follows. 'This is an experiment to find out something about the way people work out what is coming next. As you can see, there are two small lamps facing you, one with a red and the other with a yellow patch around it. One or other of these will light up briefly after each trial. Your task is to predict whether it will be the red or the yellow side which will light up. There is no guarantee that they will light up an equal number of times—one or other of them may not light up at all. All you can be sure of is that, after you have made your prediction, either the red or yellow side will light up. So now, against figure "one" at the top of the left-hand column of your slip of paper write either an "*R*" or a "*Y*" according as to whether you expect the red or the yellow side to light up. Once you have made your choice, *do not alter it*. Simply raise your hand to indicate that you have put an "*R*" or a "*Y*" in the correct position. When everyone has done that, one or other of these lamps will light up for a short while. Then you make your second prediction, and so on, always trying to anticipate whether the red or the yellow side is going to light up in each particular trial. At first you probably won't have any idea of which to expect, and you will just have to guess, but, after a while, you may find it possible to anticipate the side correctly. Try to get as many correct as possible, but remember, all results are anonymous and there are no prizes. So work inde-

pendently, and make your choice *before* the light comes on. Are you quite clear what to do? Has everyone made their first choice?' (Inspect papers.) 'All right, here is the first trial.' (Flash light for 1 second.) 'Now make your second choice.' An informal chat prior to these instructions reassured the subjects that neither their intelligence nor their personality was being measured, and anonymity was guaranteed.

Method 2. *Verbal material.* The examiner read out a list of 12 adjectives at the rate of one every 20 seconds to small groups of four or five subjects who tried to guess the related adjectives. Of the 50 subjects, 23 came from the same source as the light-guessing experiment, the remaining 27 being undergraduates at University College London, 18 to 22 years of age.

Formal instructions were as follows.

'This is a sort of guessing game. I have a list of adjectives which I am going to read out one at a time. Each of these words provides a clue to the target word which is another adjective related to the clue in some way. When I give you the clue adjective, your job is to try to guess what the target adjective is going to be. So write down the clue adjective as soon as I give it, and beside it write down what you think the target adjective is. You will have 10 clear seconds in which to make your guess and write it down. Then I shall tell you what the target word was, and go on to the next clue. Try to guess what the target is going to be even if you are not very sure, but don't write it down *after* you have been told what it is or the whole purpose of the experiment is lost. Is there anything you are not quite clear about?'

Table 2. *Lists of synonyms and antonyms*

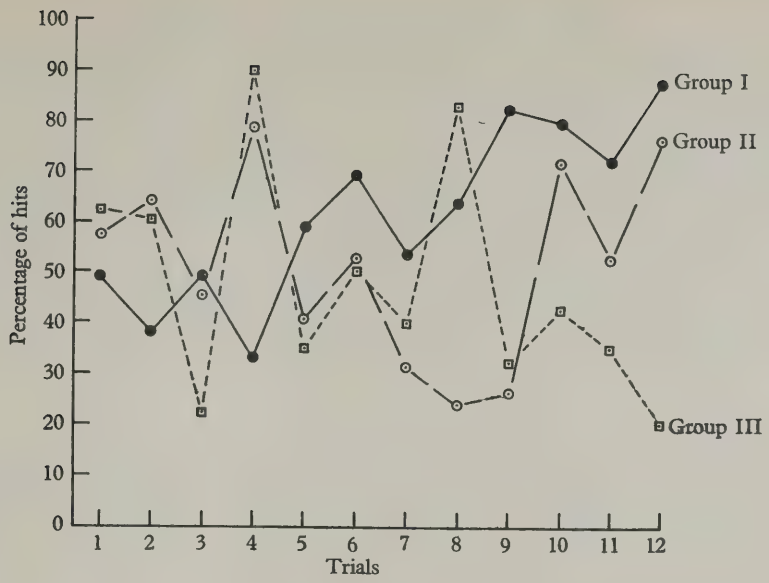
List A Target adjectives		Clue adjectives		List B Target adjectives
Fast	S	Quick	A	Slow
Broad	S	Wide	A	Narrow
Happy	S	Gay	A	Sad
Strong	A	Feeble	S	Weak
Dim	S	Dull	A	Bright
Damp	S	Wet	A	Dry
Brief	S	Short	A	Long
Good	A	Bad	S	Evil
Far	A	Near	S	Close
Low	A	High	S	Tall
Cheap	A	Expensive	S	Dear
Young	A	Old	S	Ancient

S, synonym; A, antonym

Only list structure III, as shown in Table 1, was used, counterbalanced for two groups of 25 subjects. In this case, the binary choice sequence was formed, not with two lights, but with synonyms and antonyms related to the clue adjective. The actual adjectives used are given in Table 2.

III. Results

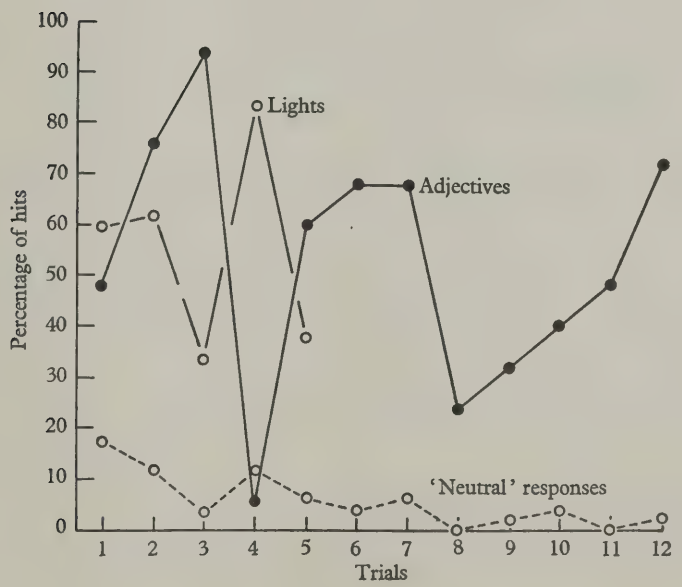
Graphical presentation of the percentage of correct responses for each of the three light-guessing groups in Fig. 1 shows the general trends under these conditions. Group I shows a fairly steady tendency to improve as the simple alternation sequence is picked up. Individual records for this group tend to reveal an abrupt shift from apparently random behaviour to insightful consistency. In the remaining two groups, our interest is mainly centred on the first five responses, which were made under identical conditions for the two groups. The first two guesses both score above chance at about the 60 % level. But having seen the same light come on twice, there is a marked (70 %) mean tendency to swing over to the other response, producing a below chance number of hits. On the next trial, which has been preceded by three instances of the one light appearing, this tendency increases to 80 % and is rewarded. The same tendency persists to a lesser degree (65 %) on the fifth trial. The question is whether, as with Humphreys's (1939) light-guessing data, reinforcement theory could generate the result, or are we committed to some sort of 'expectancy' theory?



Groups													Total S's
I	19	15	19	13	23	27	21	25	32	31	28	34	39
II	24	27	19	33	17	22	13	10	11	30	22	32	42
III	25	24	9	36	14	20	16	33	13	17	14	8	40

Actual numbers of hits

Fig. 1. 'Learning' curves for light-guessing.



Adjs.	24	38	47	3	30	34	34	12	16	20	24	36	50 S's
Lights	49	51	28	69	31	82 S's							

Actual numbers of hits

Fig. 2. Comparison of verbal and light-guessing data.

In Fig. 2 the data for the verbal material are plotted for all twelve trials, and the combined data for groups II and III previously considered are included for the first five trials to facilitate comparison.

Turning to the results for the verbal material we see very different trends emerging from the same pattern of stimuli for the first five trials. From close to a 50% level on the first trial, the tendency to choose the same type of adjective (synonym or antonym) that first appeared increases rapidly to 76% on the second trial, 94% on the third and 92% on the fourth, producing very few hits, as the change in type of adjective occurs on these trials. On the fifth trial, the strong tendency built up on the first three trials drops to the 60% level.

IV. Discussion

The results of group I are of interest for the light shed on the continuity-discontinuity question. No one who considers only the group results would deny the tendency for the scores to move from the random 50% level toward the 100% asymptote. Closer analysis of the individual results reveals a different picture. The issue to be settled is whether the subject shows any gradual improvement on his scores before he hits the correct hypothesis. This is a trickier question than it looks, and dividing the trials into blocks does not answer it. If, for instance we divide the trials into three blocks of four, then both theories predict that the fewest correct will occur in the first block and most in the third. Gestalt theory is covered by the fact that, while a few subjects will pick up the hypothesis early on, more will pick it up during the next block of trials, and so on. Only by examining the results for each individual independently can we circumvent this difficulty. If, before the subject hits the correct hypothesis, he scores above chance, then behaviourists are supported; if the subject scores around chance level they are refuted. How to decide when the subject has hit the correct hypothesis? Ideally, a verbal report should go along with each choice, describing the basis on which the choice was made. These protocols not being available, the next best solution is to adopt some arbitrary criterion in terms of sustained success. Suppose, for instance, we consider that the subject who is correct on the last five trials or more, has solved the problem. The probability of doing this by chance is 1 in 32, or about the 3% level on a one-tailed test. Any subject who does not reach this criterion we consider to have failed to solve the problem, and discard his results, confining our attention to the remaining subjects. For each subject, we consider all trials prior to the final run of hits. If he has succeeded on the last seven trials we count the number of hits and misses for the first five trials. If he gets more right than wrong we score him 'plus', vice versa 'minus', equal numbers 'neutral', and discard the 'neutral' subjects. According to the Gestalt position, the numbers of 'pluses' and 'minuses' should be roughly equal. The behaviourists predict more 'pluses'.

This analysis is set out in Table 3. Nineteen subjects reach criterion and 20 are discarded as having failed to solve the problem. Of these 19, a further 5 are discarded as 'neutral'. None of the remaining 14 scores 'plus'. This is an alarming result since neither party to the dispute makes this prediction (and the sign test shows it to be significant well beyond the 0·001 level of confidence). Something is radically wrong, and, in this case, biased sampling can be convicted as the primary offender. By considering *all* responses prior to the successful one we have ensured that at least one of these, the one immediately preceding the run is a 'minus'. This response must therefore be discarded. The effect of this

Table 3. *Analysis of pre-solution guessing behaviour of nineteen successful subjects*

Total right in final run	No. right before final run	No. wrong before final run	First sign allocation	Second sign allocation	Total wrong in initial run
10	1	1	N	+	0
10	0	2	−	−	2
10	1	1	N	+	0
9	0	3	−	−	3
9	1	2	−	N	1
8	2	2	N	+	1
8	2	2	N	+	1
8	2	2	N	+	1
8	1	3	−	−	2
7	2	3	−	N	0
7	2	3	−	N	1
7	2	3	−	N	0
*6	0	6	−	−	6*
6	2	4	−	−	0
6	2	4	−	−	0
5	1	6	−	−	4
5	3	4	−	N	0
*5	1	6	−	−	5*
5	2	5	−	−	1

* Last discards			Third sign allocation
	+	0	5
	N	5	7
	−	14	7
		19	19

is to reduce all scores in column 3 by one. Re-allocating signs on this basis there are again five 'neutral' discards, but the advantage of 'minus' over 'plus' scores drops to the 10% level of confidence (one-tailed). A difference of this order need occasion no concern, but even this slight trend is attenuated if we consider the following argument. Having been obliged to use an arbitrary criterion of five or more correct in the final run, we take the risk of including in the sample some subjects who have not solved the problem. According to the null hypothesis, these subjects have been responding on a chance basis and therefore would be expected to produce roughly equal numbers of hits and misses. These happen to be distributed so that most of the hits appear at the end of the series and most of the misses at the beginning. Hence, by including these subjects, we bias the sample in the 'minus' direction. It might further be argued that to offset this bias all subjects who get the first five or more in the series wrong should be discarded as non-random by the same token that selects the successful subjects. This procedure eliminates only two of the 19 subjects but reduces the difference between the 'minus' and 'plus' groups to 7:5 and forces the null hypothesis upon us.

So it begins to look as if the gestalt view has the better of the exchange. One fact could save at least part of the behaviourist's case. Suppose we discover from protocols that the subject sometimes verbalizes the correct hypothesis *after* he has been responding correctly for one or more trials. Then, if we base our analysis on the responses previous to the verbal report, we should find a trend in the positive direction, but only *immediately* prior to getting the correct hypothesis. No data are at present available to support this contention, but it looks a tempting interpretation. Unless and until it is confirmed by experiment there is little point in pursuing the theoretical implications of such a two-stage learning process.

When we turn to consider the data for the verbal material, several points need to be dealt with before a useful comparison can be made with the light-guessing data. There is, first, the question of whether we are entitled to regard the verbal task as a binary choice situation. No overt restrictions were placed upon the subjects beyond the hint that the adjectives would be related 'in some way', and the experimenter anticipated some difficulty when it came to marking the data, which would necessitate independent judges to assess whether a response came in the 'same', 'different' or 'neutral' category. In fact, the responses were so unambiguous that no outside judge was called in. As can be seen in Fig. 2 the number of neutral responses, even on the first trial, was only 9 out of 50, and the total in this wastepaper-basket category was 35 out of 600 (or less than 6 %) including complete failure to respond. In this sense, then, the choices were, for practical purposes, binary. The question now arises that, if the subject made a correct response with regard to type of adjective chosen, but chose the wrong specific adjective of this type (e.g. he predicted 'tall' to the clue 'short' when the specific target was 'long') was he reinforced or punished? According to S–R reinforcement theory, if the subject was reinforced, the next time 'short' or something similar appears as a stimulus there is an enhanced probability of 'tall' or something like it being given as a response. Although stimulus and response generalization is an established principle, it is not clear on what basis degrees of similarity are to be computed. The identical elements theory does not apply when dealing with meaningful material. In this case, any adjective is a similar stimulus sufficient to evoke a response that is similar in the sense of bearing a certain *relation* to the stimulus: that of being a synonym or antonym.

The beautiful learning curves that emerge are explicable only provided we accept that it is the *relation* that is being reinforced and learned. If the subject is reinforced for responding 'fast' to 'quick' then there is an increased probability that he will respond 'broad' to 'wide'. Given the same stimuli, but reinforced for saying 'slow' to the first clue, the probable response to the second becomes 'narrow'. Yet there are no grounds for supposing that 'fast' and 'broad' are more similar than 'fast' and 'narrow', or 'slow' and 'narrow' more similar than 'slow' and 'broad'. 'Fast' and 'broad' only become more similar if they are classified as synonyms to the clues. It is the total context that determines the next response, not *specific* preceding stimuli and responses. When synonyms are the stimuli, the data raise other problems for the orthodox S–R reinforcement position, as Marx & Bernstein (1955) have shown.

The data for the two light-guessing groups, taken on their own, are just as inhospitable to an orthodox S–R interpretation. Having been reinforced on the first two trials for saying 'yellow' or punished for saying 'red' there is a reduced probability of saying 'yellow' on the third trial. This paradox is even more apparent with the fourth response. To deal with this result we are obliged to talk in gestalt 'expectancy' terms or else in terms of transfer from previous learning situations, nor is any contradiction involved in talking both ways. To say that the subject expects a binary sequence to show a change after a run of two or three similar outcomes is quite compatible with the statement that the subject is prepared to respond in this way because of earlier training on binary sequences. The gestalt theorist is making a statement of present fact, to which the behaviourist predicates a genetic inference.

We are interested, though, to inquire why it is the subject does not expect a synonym after two or three antonyms or transfer his binary guessing habits to this situation. We

are further interested to know just what the earlier learning situations are that produce these non-random guessing habits. The genetic account is useful only if specific. To the first query we are bound to point out that, although synonyms and antonyms do make a binary choice situation, there is still a certain amount of variety possible within these two classes. In the light-guessing situation, the subject appears to reject the simple repeated response as too easy, whereas some sort of skill is required to find a related adjective. Whether this rejection of the simple solution is innate or acquired is another question. Perhaps the behaviourist would deal with the genetic issue by suggesting that in competitive situations an opponent varies his play as much as possible. This is particularly true of simple two-choice games played by children, such as matching pennies or picking the fist concealing a matchstick. Since these games are played only for a short while, the longer runs characteristic of a random series are inappropriate, and the short run tactic is overlearned.

References

Bitterman, M. E. (1953). Spence on the problem of patterning. *Psychol. Rev.* **60**, 123–26.

Hull, C. L. (1920). Quantitative aspects of the evolution of concepts. *Psychol. Monogr.*, no. 123.

Hull, C. L. (1952). *A Behaviour System.* Yale University Press.

Humphreys, L. G. (1939). Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. Exp. Psychol.* **25**, 294–301.

Lawrence, D. H. (1949). Acquired distinctiveness of cues. I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. Exp. Psychol.* **39**, 770–84.

Lawrence, D. H. (1950). Acquired distinctiveness of cues. II. Selective association in a constant stimulus situation. *J. Exp. Psychol.* **40**, 175–88.

Marx, M. H. & Bernstein, B. (1955). Generalisation of reinforcement among responses to synonyms. *J. Gen. Psychol.* **52**, 49–64.

McClelland, D. C. (1943). Studies in verbal discrimination learning. IV. Habit reversal after two degrees of learning. *J. Exp. Psychol.* **33**, 457–70.

Prentice, W. C. H. (1949). Continuity in human learning. *J. Exp. Psychol.* **39**, 187–94.

Spence, K. W. (1940). Continuous versus non-continuous interpretations of discrimination learning. *Psychol. Rev.* **47**, 271–88.

Spence, K. W. (1952). Mathematical formulations of learning phenomena. *Psychol. Rev.* **59**, 152–60.

Walk, R. D. (1952). Effect of discrimination reversal on human discrimination learning. *J. Exp. Psychol.* **44**, 410–19.

(*Manuscript received* 9 *January* 1957)

EXPERIMENTAL DESIGN: A CLOAK FOR INTELLECTUAL STERILITY

By J. G. TAYLOR

Psychology Department, University of Cape Town

Lewin's (1931) warning against applying in psychology the Aristotelian notion that frequency of occurrence is equivalent to lawfulness fell on deaf ears. To-day Aristotelian modes of thought are as prevalent as ever, and are backed up by an elaborate structure of statistical methods and rules for experimental design. By reason of their logical tidiness these methods have the unfortunate effect of reducing (to a level below the threshold for effective action) the discontent that should be engendered by the rotten state of current psychological theory. What psychology needs is not statistical laws but scientific laws, that is, statements, expressed in general terms, that can generate specific predictions when specific values are assigned to the general terms. Such laws are unlikely to be found by statistical investigations, since they are concerned not with populations of systems but with single systems. Thus the law of gravitation is not a catalogue of the properties of solar systems, but it explains how any given solar system hangs together. It is argued that since mind occurs in an organic system its laws are likely to be laws of an ultrastable system (Ashby, 1952), and therefore not discoverable by statistical methods. Two kinds of investigation are required. The first is exemplified by Piaget's studies of the development of mind in individual children, and the second by Kohler's work with distorting spectacles. The first shows how the organism, considered as a multistable system, adapts itself by stages to its environment. The second involves a deliberate disorganization of both behaviour and perception, and enables us to observe simultaneously the processes of behavioural readaptation and the correction of perceptual distortions. It is claimed that the results are compatible with the author's theory of the behavioural basis of perception.

I. Theory and experiment

It has often been said that what is wrong with modern psychology is that there is too much theorizing and too little experiment. Superficially this appears to be true; nevertheless I propose to maintain the opposite thesis, that there is a dearth of serious theorizing and an excess of trivial experimentation. It is true that there are too many theories, but they are not theories in the sense in which the term is understood in other scientific disciplines. A genuine scientific theory is a model that seeks to account for some specific aspect of nature *within the framework of science as a whole.* Behaviour theory satisfies this requirement, but in the field of psychology* proper, the study of mental phenomena, it seems to have been completely neglected. Without the restraint that this demand imposes, psychological theories can be multiplied *ad infinitum.* There are certainly too

* In this paper I need to distinguish between the study of behaviour and the study of conscious phenomena. Originally the term 'psychology' referred to the latter study alone, and it seems highly desirable that it should continue to be so used. When I use the term psychology I wish the reader to understand that I am specifically excluding behavioural science. It is quite possible to write to-day a 'Psychology from the standpoint of a behaviourist', but it would be a book about *mind*, that is, about conscious phenomena. Watson's book with this title was a book not about mind but about behaviour. By using the word psychology in his title he created a lot of unnecessary confusion.

There is similarly a lot of confusion over the use of the word 'mind'. I shall use it to mean the states of consciousness, actual and potential, that characterize an individual. No reference is intended to any ghost in the machine, nor to any joker lurking in the cellarage, whether Freudian, Jungian or anything else, nor to any organizing principle or any other theoretical constructs.

many theories of this kind, but there is a conspicuous lack of theories that can withstand detailed logical scrutiny and involve no conflict with any of the known laws of nature. Indeed the majority of psychological theorists seem to act as if the laws of nature were totally irrelevant to their problems.

The situation is even worse than it was when Murchison published his *Psychologies of* 1925 and *Psychologies of* 1930, for whereas the contributors to those volumes did not hesitate to admit the full extent and gravity of their differences with one another, the contestants of to-day seem to have agreed to forget the basic philosophic and scientific issues that troubled their predecessors, and pretend that those issues are no longer of any importance. Consider, for example, the nativist–empiricist quarrel. At one time the two points of view were totally irreconcilable, and each side understood quite clearly that any concession to their opponents would undermine the logic of their theories. To-day they are all eclectics: there is not a nativist who will deny that there are some aspects of mind that are modified by experience, and there is hardly a single empiricist who will not concede that certain features of mind are innate. The implicit dethronement of logic is apparently a matter of no consequence!

From one point of view this extensive unanimity might be interpreted as meaning that the science of psychology is growing up, in that there is now a substantial measure of agreement among its practitioners. This is a dangerous fallacy. What has actually happened is that, on the one hand, a growing body of evidence has made it impossible to maintain a thoroughgoing nativist theory, and, on the other hand, no really rigorous empiricist theory has yet been published, so that the psychologist with empiricist leanings is constantly being confronted with phenomena that he is unable to deduce from his 'theory'. In these circumstances, neither side is able to deliver a vigorous attack on the other, and this paralysis is sometimes interpreted as meaning that the competing views are less fundamentally opposed than they were earlier thought to be. Certainly, two eclectic points of view that borrow, in somewhat different proportions, from the same mutually incompatible philosophies, are not any more opposed to each other than the hypotheses that the devil is black and that the devil is red. But the philosophies from which both have borrowed are still mutually incompatible. It is impossible for some forms of mind to be innate and others acquired, just as it is impossible for some forms of matter to be composed of electrons and other particles in motion, while other forms are made of particles that are in static equilibrium. If we choose to ignore this fundamental contradiction, dismissing it as a pseudo-problem, we thereby preclude ourselves from asking any vital questions at all.

If we do not allow ourselves to think, there is only one way in which we can create the illusion that we are still honest scientists: we must keep the stream of experimentation flowing continuously. But this creates a new dilemma. The stream of worthwhile experiments has only one genuine source, and that is hypotheses that have some pretensions to being accepted as universal laws of nature. But if we refuse to face the most fundamental problems, we cannot formulate a single hypothesis that can be included in this class. So we must hunt around desperately for a substitute, and the most promising candidate that has appeared so far is experimental design. We can now ask the most trivial questions we like, and provided we set about finding the answer in the approved fashion, using all the new tricks of experimental design, we can be reasonably certain that someone will hail our discoveries as 'important contributions to knowledge'.

My complaint against experimental design is not that it is bad in itself. It is, on the contrary, a most admirable body of doctrine which, in its rigorous self-consistency, stands in strange contrast to the chaos of psychological theory. In fact, it is this contrast that is at once its chief attraction, and the chief source of the damage it does to psychological theory. For if a psychologist can design his experiments in such a manner as to make full use of all the apparatus for testing significance, analysing variance and the like, his demand for logical tidiness is satisfied even if the resulting contribution to science is negligible. The essence of my complaint is that, by partly satisfying the investigator's demand for logical consistency, experimental design protects him from the acute discomfort which the general theoretical chaos should induce in him and thereby inhibits him from lending a hand in the job of cleansing the Augean stables of current psychological theory. The need has been urgent for a long time, but so far the cleansing stream has been little more than a trickle, and for this I put at least a part of the blame on blind faith in the efficacy of experimental design.

This is a strange outcome of procedures that ostensibly were devised for the express purpose of answering theoretical questions, and it is important that we should try to find out what has gone wrong. I suggest that part of the answer is to be found in the origin of some of the statistical techniques that are now so widely used. Consider, for example, the analysis of variance. Originally this was a technique devised to provide answers to certain practical problems in agriculture, such as deciding which of several varieties of a plant might be expected to yield the biggest crop under different sets of conditions. It was not intended as a substitute for basic scientific research, although it could be used as a tool for testing deductions from hypotheses arrived at through basic research. The farmer cannot afford to wait for the slow process of basic research to solve all his problems; he must have answers with a minimum of delay, and carefully devised experiments in which provision is made for the analysis of variance do precisely that.

A similar urgency in the Second World War led to the widespread adoption of the same techniques by applied psychologists. In fact, the urgency here was even greater, for, whereas the agricultural scientist has at his disposal a substantial body of accepted doctrine on soil chemistry and plant physiology, the psychologist has virtually no established theories on which he can build, and therefore must of necessity use methods which do not depend on basic concepts. The successful use of such methods in a wide range of researches on the selection of personnel and the like has added enormously to their prestige and made it almost impossible for a psychologist to ignore them, even if he is engaged in pure research. If he is an 'applied' psychologist there are no other tools he can use, for, ironically enough, there is no body of doctrine which everybody agrees to call psychology that is capable of being applied. The one thing an applied psychologist cannot and dare not apply is *psychology* in any of the forms in which it is at present understood. In the circumstances, the only thing he can apply is statistical method. But this does not justify the extensive use of such methods by psychologists who are engaged in pure research. The kind of answers we have a right to expect from them are not statements of probabilities relating to isolated sets of phenomena, from which we can deduce nothing concerning any other phenomena, but general laws that are applicable throughout the whole field of psychology.

II. Scientific laws and statistical laws

The demand for universality, the justice of which no psychologist will question, is commonly understood as meaning that, for a law to qualify to be called universal, it must be shown to produce the *same results* in every person. But this type of generality is precisely what has never been demonstrated in any type of psychological experiment. By this I do not mean merely that results vary quantitatively from person to person. Quantitative variation does not invalidate a general law provided that the variations are themselves accounted for by law. What is much more serious is that there is hardly a single experimental procedure that does not produce results that differ qualitatively. Clearly a law cannot claim to be universal if it predicts one type of result, and experiment shows that there are two or more qualitatively different types of result. Confronted with this uncomfortable fact, there are many psychologists who do not hesitate to conclude that psychology is by its very nature an *inexact science*. This seems to imply a belief not merely that nobody has yet found the laws that will account exactly for all types of results, but that no such laws exist in nature. But this is absurd. Mind constitutes part of the system of nature, and it must be lawful. How then are we to discover its laws? Clearly not by statistical methods, since these do not enable us to calculate exactly the results to be expected in an individual case.

An analogy may reveal what is required. If Newton had had at his disposal not a vast amount of detailed information about a single solar system but a much smaller number of facts about each of a thousand solar systems, collected by a thousand observatories, he might conceivably have developed statistical methods for organizing this material. He might have found correlations between such variables as the number of planets in the system, the average number of satellites per planet, the average distance of the planets from the sun, and the like. He would, by this means, have learned a good deal about solar systems in general, but he could not have calculated the time and place of the next eclipse of the sun, and he could not have arrived at an understanding of the laws of planetary motion. He would have learned a lot about the ways in which solar systems differ from one another, but nothing about the way in which any one of them works. For this it was necessary to know as much as possible about one system. Fortunately Newton had no alternative, and the result of his labours was the construction of a theory that survived until the advent of Einstein's theory of relativity.

If we think of each individual mind as a separate system, then it may be said that the existence of innumerable systems, all differing from each other in detail, and all accessible to observation, constitutes a serious obstacle to the construction of universal laws of mind. The temptation to compare those systems and to think of what they have in common as being lawful, and of deviations from normality as capricious or at least unexplainable in terms of law, is still very strong more than a quarter century after Lewin's (1931) forceful exposure of the futility of this Aristotelian way. of thinking. With such a wealth of material at hand it is easy to be misled into thinking that the counting of heads may reveal something that occurs universally and will therefore enable the investigator to write a 'universal' law of nature. But a real universal law has nothing to do with universality in this sense. Newton's first law of motion, for example, states that a moving body continues to move in a straight line unless deflected by some force, but the frequency of straight paths in the celestial sphere is nil. Again, his law of gravi-

tation says nothing about frequency. It simply says that any two bodies of mass M_1 and M_2, separated by a distance d, are attracted to each other by a force that is equal to GM_1M_2/d^2. Now this law of gravitation does not tell us how many planets a solar system must have, what their masses must be, or their distances from the sun, or the shapes of their orbits. What it tells us is that, given certain facts about the sun and planets, we can calculate the behaviour of the system as a whole, and of course the behaviour of any selected unit, including any that may be artificially added to the system. Similarly, the universal laws of mind, when they are discovered, will tell us nothing directly about the actual content of any specific mind. They will be formulated in such a way that, given certain facts about an individual, including facts that are not in themselves mental phenomena, we shall be able to calculate other properties of the system. A law of nature is not a catalogue of the parts that every system must have, or a classification of the parts according to the frequency of their occurrence, but an explanation of how the parts hang together in the system. Thus a law of nature can never be stated in the form of a statistical law, although statistical laws may be derived from it.

Conversely, it will not as a rule be possible to arrive at a law of nature by means of investigations that rely largely on statistical methods. I do not mean by this that these methods are of no value at all. On the contrary, they may, by revealing loose bonds between certain variables, prompt the investigator to search for the connecting links and thus lead to the discovery of the true laws of the system. This, for example, is what Eysenck is aiming at in his investigation of the dimensions of personality. Unfortunately, the majority of psychologists continue their further search by making still more statistical investigations, usually including variables that have hitherto been neglected, and in the end they are as far away from the goal as ever. The reason for this failure is not that there is anything wrong with the design of their experiments but that they have asked the wrong kind of questions. They have neglected the vital question: what makes the system hang together in any one person?

A somewhat crude analogy will help to make the point clear. Suppose that an investigator, knowing nothing about the construction of a motor car, decided to choose as his area of research the behaviour of the speedometer needle, and to this end took a series of readings in each of a hundred different models. Just to make the problem more like a real one we shall suppose that the speedometer dials are not provided with scales, but that the investigator can measure the angular deviation of the needle. Among the variables he might be expected to record are the distances of the accelerator and brake pedals from the floor, the position of the gear lever, the gradient of the road, the direction and velocity of the wind, and, of course, the speedometer reading. He takes a succession of simultaneous readings of all those variables in each car, and then proceeds to examine his data in the hope of solving the riddle of the speedometer needle. At first the material looks completely chaotic. There is no single independent variable that is functionally related to the dependent variable, and he decides to have recourse to statistical analysis. He finds negative correlations between the speedometer reading and (*a*) the distance of the accelerator pedal from the floor, and (*b*) the gradient of the road; and positive correlations with (*c*) the position of the gear lever, and (*d*) the distance of the brake pedal from the floor. He finds significant differences between the speedometer readings when the gear lever is in first, second, third and fourth positions, but the distributions overlap extensively. He now decides to record additional data, such as the weight of the car and its

consumption of petrol, but the riddle remains unsolved. Of course we know the answer. If our investigator will only take independent measurements of the speed of the car he will find that in each system (car) the speedometer reading is a function of speed, but not necessarily the same function in all systems. He will find, moreover, that he can now dispense with statistical methods and can examine each system, considered as a matrix of pointer readings representing the several recorded variables, to determine how it hangs together. He will discover that what he at first took to be evidence of arbitrariness or caprice in his data was actually an artifact arising from the simultaneous examination of pointer readings taken from a hundred different systems. He will find that the same general principles apply to all the systems, but each of them has its own specific set of parameters, with the result that, in Ashby's (1952) terminology, the lines of behaviour of all the systems are different. Continuing to use Ashby's terms, each system is regular and absolute. It is regular because whenever it starts from a given state and a primary operation is applied to it, such as an increase in the gradient of the road or a specific depression of the accelerator pedal, the system will change to another state, and always to the same state. It is absolute because this is true no matter how the given initial state was arrived at.

Whenever we find a *natural* system, as opposed to the man-made system discussed above, that has these properties, we can begin to investigate its laws, and we shall not find it necessary to employ statistical methods or design our experiments in accordance with a plan laid down beforehand by statisticians. There is never any difficulty in devising experimental tests of a good hypothesis. The real difficulty is in finding the right hypothesis.

III. The psychological system

The man-made system discussed above is a dynamic system, and, consequently, its properties cannot be discovered by observing its state at any single moment of time. It must be subjected to the widest possible range of primary operations in all possible combinations, and the state to which it moves after each operation must be recorded. It is only when we have discovered, by this means, that the system is absolute that we can begin to speculate profitably on the laws that hold it together. Similarly, every organic system is dynamic, but it has in addition the property that Ashby has called *ultrastability*. That is, it can, in response to repeated application of certain primary operations, change the internal relations of some of its variables in such a manner that those primary operations no longer tend to throw it out of balance. It equips itself, so to speak, with an automatic gear-change that responds at once to environmental changes.

This property of ultrastability has the most far-reaching consequences for psychology. It implies that the true character of the system can never be discovered by subjecting it to a series of primary operations *after* the process of internal adjustment has been completed. This was long ago appreciated by students of behaviour, and research in this field has been fruitful precisely because experiments have been designed for the purpose of revealing the steps by which the system changes in response to new conditions. In sharp contrast to this, many students of mental phenomena have shown a marked reluctance to ask how the system arrives at the state found in a mature organism. They seem to be content to examine the system as it is now, and to try to understand it by subjecting a large number of individuals to a standard set of primary operations such as tests and

analysing the results statistically. If similar methods had been adopted by students of behaviour we should certainly have understood very little to-day about the laws of learning.

If the psychological system as it exists in a mature subject is the result of a process of adjustment, two different but complementary lines of research are immediately suggested. In the first place, an attempt should be made to discover and tabulate the state of the system at different levels of development. If mind is not an independent system in the Cartesian sense but is an essential aspect of the organic system, then the property of ultrastability leads to the expectation that this type of investigation will reveal an orderly sequence of changes, although not necessarily the same sequence in all individuals, and not necessarily covering the same period of time. The difficulty here is to devise techniques of investigation that will yield reliable results, and this difficulty has been largely overcome by Piaget, whose long series of studies of the development of mind in children reveals just the kind of orderly sequence of changes that might be expected. The second line of research is experimental rather than observational. It involves the manipulation of variables in such a manner as to throw the system out of balance, temporarily destroying its stability. If the system is, as I have supposed, part of an organic system and therefore ultrastable, it should adjust itself to the experimentally altered conditions and correct the induced instability. The first experiment of this kind was performed by Stratton sixty years ago, and he obtained the clearest possible evidence of ultrastability. When he first put on his inverting spectacles the world appeared to swing in front of him in the direction of his head movement but at twice the speed. That is, the positional stability of the visual field, which is a conspicuous feature of normal perception, was destroyed. *But it was almost completely restored in the course of the experiment.* This important result has been conveniently overlooked by critics who have been mainly concerned to prove that Stratton never saw the world the right way up through his lenses. His experiment was not continued long enough to yield unequivocal evidence on this point; but this is not true of the experiments conducted by Erismann and Kohler at Innsbruck, in which stability was so completely restored that the reversed visual field was perceived the right way round.

This clear demonstration of ultrastability in the perceptual field very strongly suggests that mental phenomena are intimately related to the learning process by which the ultrastability discussed by Ashby is achieved. By this I do not mean that perceptual responses are learned in the same way as striate muscle responses. This is the transactionalist view of perception which Pastore (1956) has recently shown to be logically untenable. I prefer the more parsimonious view that perceptual stability is a direct function of the behavioural stability which the organism achieves automatically through learning. That is, mind is a secondary product of the acquisition of an appropriate set of learned responses.

IV. Experimenting on an ultrastable system

There is not, as yet, a very large body of evidence in support of this hypothesis, just because very few experiments have been directly inspired by it. However, in the course of his normal exploratory activities the scientist often brings to light facts whose significance is not immediately obvious, and one of these facts was found in an undeservedly neglected experiment by Howells (1944). The immediate purpose of the experiment was to determine whether it was possible to induce colour-tone synaesthesia artificially, and

the method employed was to present a tonal and chromatic stimulus simultaneously and to require the subject on each occasion to *name* the colour. The colours used were a saturated red and its complementary. We shall refer to them as R and G. After the first 5000 trials, saturated and unsaturated colours were presented in alternation. We shall use the symbols r and g for the unsaturated colours. Two tonal stimuli were used, differing only in pitch. We shall refer to the lower one as 1 and the higher one as 2. In each group of twenty trials the stimulus pairs (1, R) and (2, G) were presented nine times each and the pairs (1, G) and (2, R) once each, the order being randomized. The plan after the first 5000 trials is most easily described by considering groups of forty. For the even-numbered trials in such a group the plan was the same as in the first part of the experiment; for the odd-numbered trials the stimulus pairs (1, r) and (2, g) were presented nine times each and the pairs (1, g) and (2, r) once each. The subjects were paid for their services and were warned that too many mistakes might endanger their employment.

It is easy to see that in this situation the verbal response 'red' became conditioned to stimulus 1, and 'green' to stimulus 2, so that the stimulus pairs (1, G) and (2, R) ultimately tended to evoke the responses 'red' and 'green' respectively. Learning theory enables us to deduce that the number of erroneous responses to (1, g) and (2, r) and even to (1, G) and (2, R) must have increased progressively, since drive was increased as the tally of errors mounted, resulting in larger increments of habit strength each time a correct response was given to (1, R), (1, r), (2, G) or (2, g). The results reported by Howells are in fact entirely explainable in terms of learning theory.

Howells completed his excellent study by performing two experiments which he modestly describes as supplementary but which are in fact among the most important experiments in the history of psychology. I shall describe only one of them. The apparatus for producing the chromatic stimuli was modified so that the subjects themselves could alter the proportions of red and green light reaching the ground glass screen on which the colours appeared. At the beginning of each trial the apparatus was set by the experimenter so as to produce either R or G, and the subject was required to manipulate the control until the light appeared to him to be neutral. If this was done to the accompaniment of stimulus 1 the light which the subject saw as white was invariably an unsaturated green; with stimulus 2 it was an unsaturated red. That is, *the conditioning of colour naming responses to the tonal stimuli had changed the colour perceptions of the subjects.*

It may be objected that this conclusion is not justified in the absence of a control experiment in which no new responses of any kind were conditioned to the tonal stimuli. As it happens the required control was supplied earlier by Kelley (1934) who, oddly enough, described the technique he employed as that of the conditioned response. Actually, what he did was to present certain pairs of tonal and chromatic stimuli a large number of times but without asking his subjects to do anything except attend to them. Since the experimental situation did not provide for the reinforcement of any responses the subjects might make, there was no conditioning of any kind. Consequently, if the changes in colour perception reported by Howells had nothing to do with the conditioning of colour naming responses but were entirely due to the repeated conjunction of one tonal stimulus with one chromatic stimulus, exactly the same kind of results should have been obtained by Kelley. In fact his results were uniformly negative.

These two experiments, together with much corroborative evidence which I shall review elsewhere, are sufficient to justify the formulation of the hypothesis that mind is a

secondary product of the acquisition of an appropriate set of learned responses. If this hypothesis should be confirmed, as I believe it must be, it will obviously have the most far-reaching consequences. It implies, in the first place, that mind is not an independent system with laws of its own, as Descartes believed and as many modern psychologists apparently tend to believe. But, secondly, it does not imply that mind is a useless epiphenomenon, a superfluous decoration that we could easily do without. On the contrary, it has the supremely important property that it provides an immediately available index of the state of adjustment which an individual has achieved. For example, if we find a subject who is not colour blind but who reports as white a red-green mixture that has an excess of red in it, we can infallibly conclude that he has been through some process of conditioning analogous to that which Howells imposed on his subjects. And with equal certainty we can conclude that if a group of subjects are all agreed in calling the same mixture red, they have been through a process of conditioning determined not by some inquisitive experimenter but by the natural properties of the environment in which they grew up.

This idea was vaguely apprehended by Rorschach when he constructed his ink-blot test. He understood that what a subject perceives in an ink-blot is not a function of the stimulus alone but is largely determined by his personality. What he did not understand is that this personality is made up of a large number of specific behaviour tendencies acquired in the course of his life, and that, as yet, we know virtually nothing about the mental concomitants of such tendencies. The vast amount of research that has been inspired by Rorschach does very little to fill this gap, because there is seldom any attempt to trace a specific response or pattern of responses to its behavioural foundations. Rorschach testers, too, have succumbed to the appeal of experimental design.

It now becomes clear that the primary task of psychology is to trace every aspect of mind to its behavioural roots. For this purpose the ready-made recipes of experimental design will be of little value, because the kind of experiment that is required is one in which individuals are subjected over a period of time to some kind of disruptive treatment that may be expected to produce a mental change. But individuals differ widely in the repertory of habits they bring to such an experiment, so that the treatment does not necessarily produce identical results in all. A genuine scientific law must account for the exceptions as well as the rule, and this implies that a successful experiment must involve a more detailed investigation of individual cases than is customary. The experimenter cannot rest content with a demonstration that the treatment produces the expected results in *some* subjects; he cannot have a true understanding of the laws of the system unless he can show that the same laws, operating on a subject who differs from the others in some particular respect, must produce different results.

To illustrate the point I refer, with the permission of Prof. I. Kohler, to an experiment in his laboratory in which the subject wore right–left reversing spectacles for a period of 71 days. Three experiments with similar spectacles had previously been performed, two in Innsbruck and one in Cape Town, and in all three the subjects ultimately reported that the visual field was no longer reversed. This did not happen in the fourth subject, although he wore the spectacles for a longer time than the others. He did, however, achieve the same stability of the perceptual field as the others. This result appears to invalidate the hypothesis that perception is a function of learned behaviour, since the subject was behaviourally well adjusted long before the end of the experiment. He could,

for example, ride a bicycle with ease while wearing the spectacles, although he declined, after one attempt, to master the difficulties they impose on a swimmer. Why then was there no evidence of the expected ultrastability in his perceptual experience? Actually there was evidence of it. The instability of the perceptual field under rotation of the head disappeared as in other subjects; it was only with respect to right–left orientation that he differed from them. The explanation proved to be very simple. In the course of conversation the subject revealed to me that whenever he was asked whether one object was to the right or the left of another object he invariably related the positions of the objects to the blurred edges of his visual field. Since, for a given disposition of objects relative to the median plane of the head, these positions remained invariant throughout the experiment, and since the right and left edges of the field remained invariant relative to the right and left sides of the head, the subject could always report the field to be reversed. As it happened, his previous training had induced in him a sceptical attitude towards the claim that the visual dimensions right–left and up–down could be changed experimentally. Like many others who lean towards nativism he could not swallow this idea, although he had no difficulty in accepting the equally startling claim that the perceptual instability induced by reversing spectacles could be abolished. He was able to satisfy his philosophic inclinations by developing the habit of describing the positions of objects in space in terms of a reference frame defined by his own head instead of an external reference frame. Had he taken his philosophy a little more seriously he would have recognized that the other claim was equally incompatible with it, and would have discovered that the rotations of the visual field induced by movements of the head, if described in relation to the same reference frame, also remain unchanged throughout the experiment.

We may sum up this group of experiments by saying that all four subjects learned to move in the appropriate direction relative to external objects. In addition, the first three learned to give verbal descriptions of the positions of objects in conformity with their overt actions; the fourth subject found a criterion for such descriptions that was logically, though not practically, incompatible with his overt behaviour but compatible with his pre-conceived philosophic ideas. He achieved behavioural stability without endangering his philosophical stability. In the first three all the most important modes of behaviour conditioned to retinal stimulation were transformed in the same direction, and in consequence the visual world was perceived as correctly oriented. In the last subject one sector of behaviour remained unchanged and the corresponding aspect of perception likewise remained unchanged. It will be seen therefore that the behavioural theory of perception is not only not refuted by divergent results of the same treatment but receives added confirmation from them. If the treatment had induced identical behaviours in all the subjects and divergent perceptual results had still been reported, the theory would automatically collapse.

V. Conclusion

Psychological theory as it exists to-day suffers from two closely related defects. It is not general enough, and it is not integrated with the main framework of scientific doctrine. As a direct consequence there is no body of doctrine to which every psychologist can subscribe as there is in other scientific disciplines, including the study of behaviour. Behaviourists differ from one another on the interpretation of details, but at least they are agreed in regarding the field of their inquiry as falling within the system of nature,

so that they cannot, for example, postulate mechanisms that are incompatible with the laws of physics, chemistry, physiology and neurology. Psychologists, on the other hand, often give the impression that they feel themselves to be under no obligation to consider such restraints. They seem to think that as long as they say nothing about neurology they are free to spin any kind of theory that has an air of plausibility, without stopping to consider whether it is consistent with neurology or with behaviour theory.

Now it is easy enough to design an experiment to test such a theory, and it is not at all impossible to get statistically significant results that appear to support it. But there are three other conditions that must be satisfied before it can be accepted as a law of nature. (1) It must be logically consistent. That is, it must not generate deductions that are mutually contradictory. In the science of behaviour it is customary to make some effort to satisfy this requirement, but, in psychology, it is hard to find a theory that does not violate it. (2) A valid psychological theory must be consistent with the general body of scientific doctrine. No doubt every psychologist will admit the justice of this demand, but assenting to it is not equivalent to putting it into practice. There are still habits of thought derived from the time when a wide gulf separated *Geistwissenschaft* from *Naturwissenschaft*, and particularly habits that prevent psychologists from seeing mind as an integral part of an organic system having the property of ultrastability. Specifically our second demand implies that there can be no science of mind apart from the science of behaviour. (3) Finally, a valid theory must be stated in a general form. That is, it must be so formulated that when non-contradictory values are assigned to the several variables and parameters of the system it is possible to draw specific deductions that can be tested by observation or experiment.

Every one of these conditions is necessary, and together they are sufficient to guarantee that any theory of mind that satisfies them has some pretensions to being considered a universal law of nature. If we neglect one or more of them in formulating our hypotheses, we can go on producing hundreds of researches that are impeccable from the point of view of experimental design while contributing nothing to the understanding of mind.

References

Ashby, W. R. (1952). *Design for a Brain.* London: Chapman and Hall.

Howells, T. H. (1944). The experimental development of color-tone synaesthesia. *J. Exp. Psychol.* **43**, 87–103.

Kelley, E. L. (1934). An experimental attempt to produce artificial chromaesthesia by the technique of the conditioned response. *J. Exp. Psychol.* **17**, 315–41.

Lewin, K. (1931). The conflict between Aristotelian and Galileian modes of thought in contemporary psychology. *J. Gen. Psychol.* **5**, 141–77.

Pastore, N. (1956). An examination of one aspect of the thesis that perceiving is learned. *Psychol. Rev.* **65**, 309–16.

(*Manuscript received* 25 *February* 1957)

THE EFFECT OF SET ON ACCURACY OF AUDITORY PERCEPTION

By GEORGE A. TALLAND

Massachusetts General Hospital, Harvard Medical School

Two complementary experiments are reported which compare the facilitatory effects of (*a*) sets established in various ways, and (*b*) different conditions of confirmation, on accurate recognition of words presented aurally, and on systematic distortion of words incongruent with the set.

Sets, whether explicitly established by instruction, or implicitly arrived at from cues embedded in the test material, significantly increase the incidence of accurate perception. The presence of incongruent stimuli does not noticeably reduce the facilitatory effect of a set, not even when confirmation is considerably delayed. Wide spacing of confirmatory stimuli is least conducive to the maintenance of a set.

While one set is in operation, and facilitates veridical perception, another set is unlikely to develop, even though conditions for its emergence are furnished by the stimulus material. However, if an initially established and subsequently confirmed set ceases to facilitate correct perception, it will be replaced by another appropriate set; more rapidly if only some of the words presented were congruent with it, more gradually if all were confirmatory.

Explicit sets elicit the highest incidence of distortion in accordance with expectancy, more particularly if confirmation is delayed or intermittent.

In facilitating accurate perception, double sets are less effective than single sets, whether all stimuli are confirmatory or some are incongruent. The differential effect cannot be attributed to the heterogeneous composition of the stimulus material, nor is it manifested in a manner consistent with a model of reciprocal inhibition exerted by the two sets on one another. The results agree with an inhibitory theory of set according to which sets operate by raising the thresholds of perception for all that is incongruent. The relative lowering of thresholds for the congruent would thus be a secondary effect, as distinct from the hypothetical neural model which accounts for this psychological law in terms of short-circuiting of pathways.

I. Introduction

That perception is a selective process has become a commonplace of psychology. Much ingenious experimentation has gone both into demonstrating that the sensory stimuli furnished do not entirely determine what will be perceived, and into isolating those agencies which select and structure, and also distort, that which is perceived. Theories about the selective processes in perception have been formulated in terms of hypothetical neural models (Koffka, 1935; Köhler, 1929) and also of purely psychological laws. Levine, Chein & Murphy (1942), Bruner & Goodman (1947) and Postman, Bruner & McGinnies (1949) demonstrated that we are more likely to perceive that which satisfies a need or value, or more generally a motive, as Postman & Leytham (1951) later suggested. An offshoot or this trend has been the much debated theory of perceptual defence, brought into prominence by McGinnies (1949), which in turn led Howes & Solomon (1950) to lay stress on the more objective selective influences in perception, such as frequency of usage in recognition of words, Bruner & Postman (1949) having previously demonstrated that the unfamiliar and incongruous are less easily perceived. After thus mapping out the various sources of selective sets in perception, the next task was to determine their relative effectiveness. The studies of Postman & Crutchfield (1952) and of Freeman & Engler (1955) have pointed out the complexity of this problem and the need for further

research, suggesting that set—whatever its source in a given situation—and familiarity are the two major selective factors in perception, but leaving open a third place for the structural or formal properties of the stimuli.

II. Purpose of study

(i) *Dimensions of set and of confirmation*

One object of the present study was to contribute to the dimensional analysis of the selective influence of set in perception. In setting himself this task, as well as in examining the implications of the relative effects of multiple versus single sets, the author picked up some hints from Postman & Bruner's (1949) study of multiplicity of set. One dimension of set or expectancy, that of *singleness* versus *multiplicity*, was derived directly from their study. A second dimension in the present research is type of set, distinguished by the manner in which it has been established. An expectancy can be aroused *explicitly* by instruction or warning, e.g. by announcing that pictures of blue flowers will be shown: a set can also be induced *implicitly* by embedding some cue in the situation, e.g. by listing the names of a few Roman emperors (cue), and then following these up with the names of several others (test series). Sets thus established are maintained if they are corroborated, if they help towards accurate or structured perception, or if they satisfy some need. Inappropriate, i.e. unconfirmed, sets also operate selectively until they are relinquished. Needs being as loosely defined as they are, no set can be shown to be motivationally inappropriate. However, the maintenance of a maladaptive, e.g. a paranoid, set is a problem of abnormal behaviour, and need not concern us here. In contrast to appropriate sets, i.e. expectancies confirmed by subsequent events, a set will be regarded as inappropriate when it creates an expectancy of events which will not ensue. More specifically, it will be called a *counter set* when these events could be anticipated by an appropriate set, the emergence of which it prevents or delays. The second dimension, therefore, ranges from *explicit set*, through *implicit set* and no set to *counter set*.

The third dimension is defined by the degree of confirmation. The theoretical extreme values are those of *complete confirmation*, i.e. every event corroborating the expectancy, and complete absence of confirmation. In this study, besides complete confirmation, an intermediate instance will be considered, namely, *partial confirmation* which means that as many instances fail to confirm as do confirm any one set in operation. These instances can be inserted at different rates, so that confirmation is either fairly evenly *distributed*, or widely *spaced* at first, with the bulk presented in one block at the end of the series, or altogether *delayed*, so that all instances of disconfirmation precede the first stimulus congruent with expectancy. Thus we have three variants of partial confirmation.

In accordance with this dimensional model, one object of the study was to determine the relative facilitatory effect of expectancy on accuracy of perception, varying sets along the three dimensions outlined. To some extent, this task amounted to extending, from the visual to the auditory modality, the work done by other experimenters.

(ii) *The mechanism of set*

There are several explanations of how sets or selective factors in perception, whatever they be called, operate. Obviously, some facilitatory process is involved, phenotypically. In a hypothetical model of the nervous system, however, this can be conceived either as

a central mechanism which raises an otherwise subliminal excitation above threshold level, as suggested by Hebb (1949), or as the effective outcome of an inhibitory process which excludes all or some interference, a view put forward by Miller *et al.* (1951). Postman & Bruner (1949), analysing the comparatively lower facilitatory effect of multiple sets, advanced a neurological model of reciprocal inhibition by competing responses. Resorting to a different model, they also suggested the analogy of 'sharpness of tuning' which is maximal with a single set and most blurred when the organism's receptivity is indiscriminate.

Now, there is no reason why a set should tune less sharply in conjunction with another set, if all that happens is the putting into operation of a central neural process. The appropriateness of the argument of reciprocal inhibition must be demonstrated empirically, as must any explanation in terms of interference from irrelevant or incongruent material. There is no *a priori* reason for assuming that the capacity of the organism, either at the receptor or at subsequent stages of perception, is so limited as to accommodate only a single set at maximum efficiency. This explanation, based on work with motor sets, is not necessarily applicable to expectancies established within the boundaries of a concept (which may be as wide as names of famous men, or as limited as South American states). Experiments comparing single and multiple sets have used expectancies defined by a category or concept, a generic term which facilitates the perception of the specific instance. These categories are pre-established, as is, for example, that for blue flowers. The tuning of the set becomes blurred when we add white and purple flowers to the blue, because we have no acquired habit of response to 'blue, white and purple' as against all other colours. By setting into operation the mechanism which lowers thresholds for white, the independent mechanism controlling the threshold for blue need not be affected. On the other hand, if a set operates selectively by excluding or raising the threshold for the irrelevant, then the boundary between congruent and incongruent becomes not only wider after adding white and purple to the blue, but also less clearly defined, and may indeed cease to be a continuous boundary.

A second object of this study was to test the theories of interference and of reciprocal inhibition, put forward to explain the relative inefficiency of multiple sets as selective factors in perception. Further to this, the results of the experiments were analysed in relation to the alternative models of facilitation by lowering thresholds for the congruent, or by raising thresholds for the incongruent.

III. METHOD

(i) *Subjects*

All subjects were adults, members of the medical, psychological and clerical personnel and of several training programmes at the Massachusetts General Hospital, Boston. Altogether 114 subjects participated in this study; 36 in the control series, 54 in Exp. I, and 24 in Exp. II.

(ii) *Procedure*

Conditions for the utilization of mental sets were provided by presenting ambiguous auditory stimuli for recognition. The effect of set was tested by its facilitation of correct vocal reproduction of words presented aurally at the threshold level of intelligibility. All test lists and instructions were recorded and presented on tape by the same voice and at

a uniform level of intensity (except that this level was considerably louder for all instructions). Each subject's individual threshold of intelligibility was determined by means of a pilot list of 24 mixed words which were not drawn from any one category. The subject's task was to repeat these words as he heard them, and the experimenter set the volume and tone controls so that about as many words were reproduced correctly as incorrectly or not at all.

The subject was seated at a distance of about 4 feet from the tape recorder. The experimenter told him that his task would be to repeat aloud words spoken on the tape recorder, and, if necessary, also assured him that this would not be a projective test. The experimenter further informed the subject that all instructions were recorded. The first of these read: 'You will be hearing several lists of words. I shall read them just loud enough for you to hear me. Therefore you may not clearly understand *every* word you hear. Don't let that disturb you. Repeat out loud every word as it *sounds* to you.' This instruction, which always preceded the pilot list, differed in the *control series* to the extent of announcing 'a list' instead of 'several lists'. In the two experiments, each test list was introduced by an appropriate instruction, e.g. 'Now you will hear names of animals', or 'This time you will hear names of birds and names of parts of the body'. When the experiment called for an *implicit set*, the instruction was: 'Now you will hear another list of words.' These instructions were clearly intelligible, no matter how low the volume control was after calibration. The silent interval after each word was long enough for the subject to repeat it aloud three times at the same pace.

The *control series* consisted of one list of 108 words; the two experiments were divided into several test lists. These lists varied in length according to what type of set and confirmation they tested. The basic unit was an array of 12 words, the number of instances presented in confirmation of any one set, and also the number of mixed incongruent words providing for partial confirmation. These mixed words (no two of which, within any array of 12, represented the same conceptual category) will be referred to as *padding*. Arrays confirming sets were drawn from the following four classes: animals, birds, parts of the body, names of food. These will be referred to by their initials: *A*, *B*, *P* and *F*, and the symbol of padding will be *X*.

Test lists for a single set consisted of 12 words drawn from one class, if confirmation was complete, or of 24 words (half of which were padding), if confirmation was partial. Double sets were tested by 24 words, drawn in equal proportion from two classes, providing for complete confirmation; partial confirmation required adding another 12 words of padding. *Implicit set* was induced by a prelist of 6 words drawn from the same class as the test list. No other indication was given that the latter would consist entirely or partly of words representing the same category, or indeed that they would be derived from any one category. *Counter set* was tested immediately after a list had confirmed an *explicit* set, without break or pause other than that between any two words. The test list for counter set was either an array of 12 words drawn from one class, but one different from that for which the subjects was set, or a list of 24 words, half of which were drawn from one class other than expected, the other half mixed. Test lists derived from more than one category were so composed that no more than two words representing the same class, nor more than two padding words should run in sequence. The selection and ordering of words for each test list were based on the findings of the *control series*.

Responses were recorded under four headings: correct; none; incorrect word or

meaningless sound; incorrect word which is congruent with a set in operation. This last class, which Allport (1955) described as 'hypothesis-confirmation by stimulus-deformation', was here called *assimilation*, i.e. assimilation of the stimulus input to expectancy. The number of correct reports within each array of 12 words was scored separately for each subject. Instances of assimilation were summated for each experimental condition across the entire sample of subjects and converted into percentage of total errors, but not averaged for tests. Differences reaching the 1% probability level were considered significant.

IV. The control series

The design of the experiment called for 36 words from each of the four conceptual classes and for 60 mixed words. The latter were so selected as to match the classified words they were to be combined with, both in respect of Thorndike & Lorge (1944) frequency count and, as far as possible, in phonetic structure. Since some pilot runs had indicated that word count rating was entirely unreliable for predicting relative auditory intelligibility, this property of the test material had to be determined empirically. For this purpose three control series were run.

Each of these was presented to 12 subjects, and consisted of the same 60 mixed words (*X*), and 12 words from each class (*A*, *B*, *P*, *F*). The order of these words was random, except that those representing any one category were evenly spaced, since, in this instance, the establishment of a set was to be avoided. For the same reason only a third of the words from each class was presented to any one subject. By counting the number of correct reports, an empirical rating of intelligibility was assigned to each classified and padding word. On this basis, three matched arrays were compiled within each category so that each contained about the same number of words with highest rating, second highest, etc.; and these arrays were also equated for order. Five arrays of padding were constructed on the same principle.

Altogether, 1231 correct responses were made to the 60 mixed words by 36 subjects, as compared with 898 to the 144 classified words. The former, of course, were presented three times as often, and therefore these results must be viewed in relation to possible totals of 2160 and 1728. In order to render the results comparable with certain findings of the main experiments, they were converted into correct reports out of lists of 12 words, to give means of 6·84 and 6·24. The difference is not significant ($t = 1{\cdot}58$) and, in view of its direction, no allowance need be made for differential intelligibility in testing the effect of set, since the words for which set was established in the two experiments were not inherently more intelligible than those used for padding.

The ranges of correct responses were 5–35 for mixed, 2–11 for classified words, out of possible totals of 36 and 12 respectively. With means of 6·61 (*A*), 6·11 (*B*), 5·78 (*P*), and 6·44 (*F*), the four classes did not represent different populations of intelligibility, the *F* ratio being 0·52.

V. Experiment I

(*i*) *Design*

The purpose of this experiment was to compare the facilitatory effects of sets on correct perception in terms of the three dimensions of the present study. This required setting up matched conditions for testing: (1) explicit, implicit and counter set with complete confirmation; (2) explicit, implicit and counter set with partial confirmation; (3) single set and double set with complete confirmation; (4) single set and double set with partial confirmation. Complete and partial confirmation, *per se*, were compared only as

related to double set, in this experiment. The conditions of comparison demanded identical test material, i.e. lists drawn from the same class of words. One category was used for each of the first two tests: A for (1) and F for (2); tests (3) and (4) shared the remaining two categories, B and P. These last two tests also had in common tests of double set with complete and with partial confirmation, so that (4) provided data for comparing the interference effects of padding with those of a second set. In both tests (3) and (4) only explicit sets were in operation. Each of the four tests of comparisons entailed three tests of differences between means. Although the three arrays were matched within each category, possible errors arising from differential intelligibility were eliminated by presenting each array, with its associated padding or second array if it had one, under all three conditions compared. For this reason three tapes were recorded so that, for example, the twelve animal names for which an explicit set was established on the first tape, were listed on the second tape so as to test a counter set, and were presented with an implicit set on the third. All test tapes were re-recorded from a master tape. The design of Exp. I is set out in Table 1, with the following code. The letters A, B, P, F, symbolize the four classes of names, X the mixed words or padding; suffixes stand for first, second, third array within each class, and for the five arrays of padding; a line above the code letter (e.g. $\bar{F}$) indicates explicit set; a dot above the letter (e.g. $\dot{A}$) indicates implicit set; and an asterisk (e.g. $\bar{P}*A$) indicates the introduction of a counter set.

Table 1. *Design of Experiment I*

Tape I	Tape II	Tape III
$\dot{A}_1$	$\dot{A}_2$	$\dot{A}_3$
$\bar{B}_1X_4*\ F_1X_1$	$\bar{B}_2X_4*\ F_2X_2$	$\bar{B}_3X_4*\ F_3X_3$
$\bar{F}_3X_3$	$\bar{F}_1X_1$	$\bar{F}_2X_2$
$\bar{B}_3\bar{P}_3X_5$	$\bar{B}_1\bar{P}_1X_5$	$\bar{B}_2\bar{P}_2X_5$
$\bar{A}_3$	$\bar{A}_1$	$\bar{A}_2$
$\bar{B}_2\bar{P}_2$	$\bar{B}_3\bar{P}_3$	$\bar{B}_1\bar{P}_1$
$\dot{F}_2X_2$	$\dot{F}_3X_3$	$\dot{F}_1X_1$
$\bar{P}_1*\ A_2$	$\bar{P}_2*\ A_3$	$\bar{P}_3*\ A_1$

Each tape was divided into three sections of approximately equal length, as shown by the broken lines. Since the experiment consisted of eight test lists, the influence of fatigue, of adaptation and, more particularly, of test sophistication had to be considered. It will be recalled that instructions were not completely confirmed in every instance; partial confirmation also implies disconfirmation, and that with massed effect in instances of counter set. It is possible that, after these experiences of disconfirmation, the effect of a set would be different from what it was before the first encounter of padding or counter set. In order to eliminate systematic errors arising from this source, successive subjects began the experiment with different sections; one in the order of recording, the next with § II and finishing with § 1, and the third with § III first and § II last. The final design thus became three-factorial, testing differences between (1) the experimental situations compared; (2) the three tapes; and (3) order of presentation. Subjects were assigned in serial order to the nine cells of the design, first going through the three starting positions of Tape I, so that each cell contained a total of six.

(ii) *Results*

The principal results of Exp. I are set out in Table 2. In accordance with the experimental design, four analyses of variance were calculated. The criteria of homogeneity of variance were satisfied in every instance but one, the exception being the test of differences due to different single sets with complete confirmation ($\bar{A}$, $\dot{A}$, $*A$), in which scores under counter set were not representative of the same population as those under explicit and implicit setting. With regard to order of presentation and tape, the variances were homogeneous, and therefore analysis of variance was calculated with only two factors considered. Differences between counter set on the one hand and explicit and implicit set on the other were tested by Wilcoxon's T for unpaired replicates (1949), that between the latter two conditions by Student's t. The last of these tests gave a ratio of 2·46 ($P > 0{\cdot}01$); both T values (2237 and 2515) were below the 1% limit of 2532.

Analysis of variance showed no significant differences due to order of presentation or to tape in this (or in any of the other) test except that between single set with complete confirmation and double set with partial as well as complete confirmation (P). In this analysis, scores on the three tapes differed ($F = 5{\cdot}67$, D.F. $= 2, 135$),† and it also showed the only instance of a significant interaction effect ($F = 4{\cdot}29$, D.F. $= 4, 135$), which was a product of tape and order of presentation. Much significance, however, cannot be attributed to these results because the variances of scores on the three tapes were not homogeneous.

Table 2. *Mean correct response within arrays; total assimilation; and assimilation as proportion of total error, under the various conditions of set and confirmation in Exp. I*

		Mean correct response	Assimilation	Assimilation / Error
Classified words				
Single set, explicit				
Complete confirmation	$\bar{A}$	9·24	26	0·174
	$\bar{P}$	9·54	43	0·323
Single set, explicit				
Partial confirmation	$\bar{B}$	9·37	20	0·140
	$\bar{F}$	8·52	35	0·188
Single set, implicit				
Complete confirmation	$\dot{A}$	8·55	34	0·183
Partial confirmation	$\dot{F}$	8·17	22	0·106
Double set, explicit				
Complete confirmation	$\bar{B}$	8·17	23	0·111
	$\bar{P}$	8·06	39	0·183
Double set, explicit				
Partial confirmation	$\bar{B}$	7·78	29	0·128
	$\bar{P}$	7·46	29	0·118
Counter set				
Complete confirmation	$*A$	7·02	67	0·249
Partial confirmation	$*F$	7·02	22	0·082
Mixed words				
With explicit single set	$X_{\bar{B}}$	3·65	31	0·169
	$X_{\bar{F}}$	3·78	70	0·157
With implicit single set	$X_{\dot{F}}$	4·80	49	0·118
With counter set	$X*_{F}$	5·55	28	0·082
With double set	X_{BC}	5·48	26	0·071

Differences resulting from type of set, the principal experimental variable, were significant in each of the three-factorial analyses. With D.F. $= 2, 135$, F ratios were: 14·55 between explicit, implicit and counter set with partial confirmation (F); 23·03 between single set and double set with complete and the same with partial confirmation (P); and 6·60 between single set and double set with complete and with partial confirmation (B).

Breaking down further differences within each of the four triads, the following were found significant: explicit or implicit set versus counter set, whether confirmation is complete or partial; single set with complete or partial confirmation as against double set with complete or partial confirmation. In neither instance does the introduction of padding significantly diminish the effect of a double set, nor is the facilitatory influence of an explicit set significantly larger than that of an implicit set. The differences between these two, when confirmation is complete ($\bar{A}$, $\dot{A}$), is significant at the 5% level ($F = 5{\cdot}41$, D.F. $= 1, 50$) but not significant ($F = 0{\cdot}78$) when confirmation is partial ($\bar{F}$, $\dot{F}$).

The means listed in pairs under the first six headings of Table 2 are, in no instance, significantly different one from the other. Under conditions of double set, individual

† D.F. is here expressed by 2 numbers. The first refers to the number of comparisons being made; the second to the total number of instances involved.

subjects tended to recognize about as many items from one category as from the other. With complete confirmation, 14 subjects correctly reproduced just as many names of birds as of parts of the body; 19 reported more of the former, 21 more of the latter, with differential totals of 47 and 41. The mean difference between the two categories for the entire sample was 1·66; with a S.D. of 1·26. With partial confirmation, eight subjects gave the same number of correct responses to birds and to parts of the body, 26 reported more of the former, 20 of the latter, with differential totals of 52 and 41. Mean difference was 1·72 with a S.D. of 1·48.

Testing the effects of the various sets on correct reproduction against the results of the control series, these were found to be significant in every instance. Whether explicit or implicit, single or double, a set makes for a greater number of accurate responses, than does no set; a counter set results in significantly fewer correct responses. Correct report of mixed words listed as padding was significantly less frequent in every instance than correct report of the same words when presented without a set, i.e. in the control series. These scores also differed significantly according to type of set, more particularly between mixed words presented with explicit, implicit and counter set ($X_{\bar{P}}$, X_P, $X_{\dot{P}}$) and between single as compared with double set with complete confirmation ($X_{\bar{P}}$, $X_{\overline{BP}}$), the F ratios being 8·57 (D.F. = 2, 159) and 15·34 (D.F. = 1, 106); the differences in the inhibitory effect of sets faithfully mirror their facilitatory effects, even to the extent of a difference significant only at the 5% level between explicit and implicit single set ($F = 4{\cdot}85$, D.F. = 1, 106).

Each of the five mixed arrays elicited significantly fewer correct responses than its associated classified array of names. Taking the results of the entire sample of subjects, perfect negative correlation obtains between correct reproduction from the classified arrays and their associated mixed words. Summating the two, the means of combined lists range between 6·15 and 6·51, showing no significant differences.

The errors divide about evenly between no response and incorrect response, the latter including instances of assimilation. The 54 subjects refrained from responding to the 144 classified names in 1206 instances and to the 60 mixed words in 978 instances; the corresponding totals for incorrect reports are 1230 and 1006, respectively. The over-all ratio is 494 no responses to 506 incorrect calls, and this varies between 583 to 417 when there is a single explicit set and complete confirmation, and 411 and 589 for mixed words with a double set. There is no consistent relationship between type of set and direction of this ratio which appears to be rather (though by no means predominantly) a function of the words listed.

When a double set was in operation more errors were made by no response under conditions of complete confirmation, and by incorrect calls under partial confirmation. The total incidence of assimilation, however, was slightly larger in the former situation. Table 2 gives details of assimilation from classified arrays; in the case of double set, however, one must also examine the categories to which responses assimilate. Incorrect reports of names of birds are listed as assimilation irrespective of whether they represent names of other birds or of parts of the body, and the incidence of these two is not necessarily equal in frequency. The totals under conditions of complete confirmation were 26 assimilation to birds, 36 to parts of the body; and under partial confirmation, 25 and 33, respectively. In neither instance is the difference significant between assimilation to or from one or the other category.

Comparing incidences of assimilation in proportion of total errors, i.e. in relation to their potential, significant differences were found in several tests. An explicit single set facilitated significantly more assimilation both to and from class *P* than did a double set, whether confirmation of the latter was complete or partial. Under conditions of complete confirmation, 39 words of this class, out of a total of 213 errors, were misreported congruently with one of the sets in operation (*B* or *P*); 36 words were misreported as parts of the body, out of a total of 420. With $\chi^2 = 6{\cdot}82$ and 8·34 (D.F. = 1), both these frequencies are significantly lower ($P < 0{\cdot}01$) than the 43 errors out of 133 faulty responses when parts of the body were presented with an explicit single set.

The totals of 67 and 22 instances of assimilation under counter set combine 'stimulus-deformation' in accordance with the set originally established and confirmed, and with the other set tacitly introduced while the first was still in operation. Between these two, the entry of 67 for complete confirmation ($\bar{P}*A$) divides into 45 food and 22 animal names. For partial confirmation ($\bar{B}X*FX$), over half the instances of assimilation both from classified and mixed words are congruent with the newly introduced set: 13 out of 22 food words, 17 out of 28 mixed words. With a counter set in operation and under conditions of complete confirmation, the median rank of assimilation in accordance with the previous set is 4·8 as against 8·9 for the new set. The corresponding ranks under conditions of partial confirmation are 10·7 and 7·7. This reversal of the positions is notable, even though it may be an artifact of the lists of words which were not identical under the two conditions. It agrees with another finding, namely, that the establishment of a new set contrary to one previously adopted is more gradual and also more definite when confirmation is continuous. Under these conditions, and with a counter set in operation, a division of correct responses into two or three successive phases shows a consistent upward gradient. The increase from 164 to 217 or from 93 to 137 and finally to 151, is significant as tested by χ^2 (7·51, D.F. = 1; and 16·96, D.F. = 2). When both original and counter set are only partially confirmed, this trend is far less consistent.

VI. EXPERIMENT II

(i) *Design*

The purpose of this experiment was threefold: (1) to retest a finding based on Exp. I, i.e. that complete confirmation does not make for significantly more correct responses than partial confirmation; (2) to explore possible differential effects of set as a function of the dispersion of partial confirmation, whether it is *distributed*, as in Exp. I, *delayed*, or *spaced* in relatively wide intervals, with the bulk delayed; (3) to discover whether, in addition to an explicit set, another set does emerge implicitly, if suggested by and appropriate to the padding words.

In Exp. I, different arrays of words were used to test the effect of set with complete as opposed to partial confirmation. In this experiment, both conditions of confirmation were tested with sets for animal names, arrays A_1 and A_3. The duplication entailed by this design provided data to test whether the samples of subjects participating in the two experiments represented the same population in terms of correct recognition of the test words. Food word arrays F_1 and F_3 were used to test the different models of partial confirmation; list B_2P_2 to explore the emergence of a second set. Array F_1 was combined with padding X_1, F_3 with X_3; partial confirmation of set for animal words was *distributed* among items of mixed word list X_2. *Spaced* confirmation (distinguished by suffix '*s*') meant that words congruent with the set occurred in positions 1, 3, 7, 14 and 17 to 24 in a list of 24. *Delayed* confirmation (the code letter of which is suffix '*d*') followed *en bloc* the entire array of 12 padding words. All sets for animal and food names were given explicitly.

Two tapes were recorded. Following the pilot list, the order of test lists was: set for animal names, complete confirmation ($\bar{A}$); set for food names, delayed confirmation ($\bar{F}_dX$); set either for birds or for

parts of the body, a list containing 12 of each (B/P); set for food names, spaced confirmation ($\overline{F}_s X$); and set for animal names, distributed confirmation ($\overline{A}X$). On one tape, A_1 was confirmed completely and A_3 partially; on the other tape, the design was reversed. Similarly, with F_1 and F_3, each being spaced on one and delayed on the other tape. On one tape test list B/P followed set for birds, and on the other a set for parts of the body. Using the symbols, the two tapes were constructed in the following order. Tape I: $\overline{A}_1$; $\overline{F}_{d1}X_1$; $\overline{B/P}$; $\overline{F}_{s3}X_3$; $\overline{A}_3X_2$. Tape II: $\overline{A}_3$; $\overline{F}_{d3}X_3$; $\overline{B/P}$; $\overline{F}_{s1}X_1$; $\overline{A}_1X_2$. Six subjects were tested on each tape in the original, six in the reverse order.

(ii) *Results*

The results of this experiment are shown in Table 3. Tested by mean correct report of animal names, the sample represents the same population from which subjects for Exp. I were drawn ($t=0{\cdot}35$). There are no significant differences due to order of presentation in any of the lists, between the two arrays of animal or the two arrays of food words, between the effects of set for birds or for parts of the body. Furthermore, no significant differences were found between correct reproduction of animal words, whether set for these was completely or partially confirmed ($t=1{\cdot}37$). Differences in correct repetition of classified

Table 3. *Mean correct response within arrays; total assimilation; and assimilation as proportion of total error, under the conditions of Exp. II*

		Mean correct report	Assimilation	Assimilation / Error
Explicit single set				
Classified words				
Complete confirmation	$\overline{A}$	9·25	10	0·182
Partial confirmation	$\overline{A}$	8·62	11	0·136
Delayed confirmation	$\overline{F}_d$	8·21	24	0·263
Spaced confirmation	$\overline{F}_s$	7·54	22	0·206
Partial confirmation	$\overline{B/P}$	9·37	8	0·127
Padding				
Mixed words				
Confirmation delayed	X_{Fd}	2·92	51	0·234
Confirmation spaced	X_{Fs}	3·54	34	0·167
Confirmation distributed	X_A	3·71	19	0·095
Classified words				
Confirmation distributed	B/P	3·21	20	0·095

words are not significant, whether confirmation is *delayed*, *spaced* or *distributed* as in $\overline{F}$ of Exp. I, the F ratio being 2·92 (D.F. = 2, 99). The difference between *spaced* and *distributed* confirmation, however, is significant at the 5% level ($t=2{\cdot}37$, $n=77$) and significant beyond the 1% level ($t=4{\cdot}11$) if we compare the first 10 food words presented in either distribution. In no other instance would a similar reduction of the arrays compared yield a significant difference, if complete arrays of 12 did not. Correct report of expected words in a mixed series of birds and parts of the body (B and P) give a mean of 9·37 which is precisely the same as correct reproduction of bird names distributed among padding words. This mean is also significantly higher ($t=8{\cdot}66$ and 9·56, D.F. = 77) than that of either category if presented with a double set. Correct report of classified names, presented with a set for a different category, is certainly no more frequent than of mixed words under similar conditions, nor is it significantly smaller. Assimilation occurs about as commonly under conditions of either experiment, and does not vary with the spacing of confirmation.

VII. DISCUSSION OF RESULTS

Using a different method of presentation, and measuring the effect of expectancy by frequency of success rather than by degree of ambiguity, this experiment has confirmed previous findings concerning the facilitatory influence of sets. Both the empirically most conclusive and theoretically most suggestive results were found in a situation in which more than a single set was appropriate either simultaneously or successively.

Sets emerging in the process of performance facilitate accurate perception as effectively as those expressly introduced so as to structure the situation in advance. Their inhibitory effect on incongruent stimuli is also about equally strong. These implicit sets are possibly not as powerful as explicit sets in distorting the actual input so as to fit into their terms of reference; this is the case, in particular, when some of the stimuli presented quite clearly do not fit, i.e. padding words. The present experiment has not succeeded in exploring the process in the course of which such implicit sets emerge. The cue series, by which these sets were induced, consisted of six words, each of either high or medium rating on basic intelligibility. Therefore, its reduction to five, to four or to even fewer items, does not substantially change the results, even if one considers comparing scores based on only partly identical test material. Further work will be necessary to determine the boundaries along which implicit sets emerge. They certainly do come into operation under conditions like those of this experiment; their development is delayed if previously another set had been operating, and the longer delayed the more consistently that set was confirmed; continued confirmation of such an earlier set, even though only intermittent, utterly impedes the emergence of an implicit set.

Sets are relatively persistent dispositions. The experimental conditions of the present study did not probe their limits, but they have shown that sets are maintained even though confirmation is considerably delayed. There is some evidence that initially loose spacing followed by solid ranks of confirmation is less conducive to the maintenance of expectancy than is evenly distributed or continuous or delayed mass confirmation.

The amount of padding (or incongruent) stimuli used in this study with single and double sets did not effect disconfirmation. It served to demonstrate the interference effect of sets on perception of incongruent material; these effects tend to be proportionate to the facilitatory effect on congruent sensory input, so that the amount of information transmitted, from relevant and irrelevant stimuli combined, is about constant under different conditions of expectancy. This rule, however, only holds in global terms, i.e. for the entire sample of subjects as a whole, not for individual subjects.

Like Postman & Bruner's (1949) tachistoscopic study, this auditory experiment has also concluded that the effect of expectancy is weaker with a double than with a single set. The present study also agrees with theirs in finding that sets, when multiple, exert a reduced effect by assimilation or formulation of congruent 'hypotheses', However, it has not traced a consistent tendency to abstain from responding rather than to report incorrect words under conditions of partial as well as complete confirmation. When incongruent stimuli are included among the confirmatory, the effect of a double set is, naturally, less pronounced than that of a single set, but not significantly less so than that of a double expectancy with complete confirmation. This conclusion corresponds to the finding that the presence of incongruent material does not, of itself, alter the effect of a single set. Since the facilitatory effect of a double set is less than that of a single set

under conditions of partial confirmation, this decrement cannot be attributed to the heterogeneity of the stimuli.

The results fit well into Postman & Bruner's 'hypothesis' theory, and also agree with their view that sets operate by excluding the incongruent as well as by lowering thresholds for the perception of the congruent. In fact, the findings agree best with a model that treats the latter effect as a function of the former which can be conceived as an inhibitory process. That, operationally, sets do exercise an inhibitory effect is clearly demonstrated by the findings that (*a*) odd, i.e. unclassified or mixed, words are less frequently reproduced when incongruent with a set than they are when presented while no set is in operation, and (*b*) once a set has been established and continues to facilitate perception, another set will not emerge spontaneously, even though cues are given far in excess of what would be needed if no other set were in operation, and even though two sets can be effective simultaneously. Of course, failure to report is not the same as failure to perceive. However, our subjects had been instructed to repeat whatever they heard, and the fact that they made rather more errors by incorrect calls than by no response suggests that they observed this rule.

The experimental evidence here presented merely adds to the massive documentation of the inhibitory effect sets exert on that which is incongruent with their terms of reference, that is, of inhibition in an operational sense. Is it this same inhibitory effect which accounts for their relatively lower facilitatory influence, when presented in double or multiple conjunction? Postman & Bruner consider inhibition primarily in this context of rivalry and tentatively go beyond a purely operational definition of the term by hinting at a neurological parallel of responses competing for a common pathway. The same model may, but need not, be extended to account for the effect of a single or multiple set, as against all that for which no set is in operation. The findings of this study are presented in support of this extension of the inhibitory model, but not specifically in support of the analogy of reciprocal inhibition. In accordance with the latter, either of two response patterns to double sets would be expected: (*a*) one of the simultaneous sets becomes prepotent and operates with the same effect as a single set, while the others cease to have a facilitatory influence on perception; (*b*) the two or more concurrent sets each dominate a phase at the end of which they submerge for the duration of another phase or until newly evoked. The former, i.e. the adoption of only one out of two concurrent sets, is characteristic of some psychotic patients we have experimented with. It did not, however, characterize the subjects of this experiment; these subjects accurately reproduced about equal numbers of words from each of the rival categories under a double set, and did this in frequent alternation rather than in longer phases. Lastly, a model of reciprocal inhibition could not account for the increase in correct recognition of irrelevant words under double as against single set, unless it also implied reciprocal inhibition of inhibitory effect.

What reason is there for hypothesizing that sets facilitate perception by virtue of an inhibitory process, by raising thresholds for perceiving the incongruent, rather than by lowering them for the congruent or by summating the peripheral excitation with a central process? If sets operated by this latter mechanism, the introduction of a double set should not result in diminished facilitation, unless we allow for reciprocal inhibition or attribute a weakening of the expectancy effect to interference from incongruent stimulus material. Both these explanations have already been ruled out in the previous discussion. According

to this hypothesis, moreover, an increase of assimilation would be expected from single to double set, since a larger selection of responses—words congruent with each of the multiple sets, as against those drawn from one only—can be elicited without the necessity of confirmation by the sensory input. This is precisely what occurred in the experiment with counter set. No such rise in the incidence of assimilation occurs with a double set; in fact it falls below the level characteristic of a single set. In view of the wider scope offered for assimilation by the smaller total of correct responses under a double set, this result is the more significant. It is difficult to explain in terms of duplication of lower thresholds or of other facilitatory central processes, unless by hypothesizing that the pathways activated in these processes are so limited that they cannot subserve selective behaviour directed toward two or more previously discrete conceptual categories. A brain model of functional equipotentiality or of alternative circuits does not warrant such a structural restriction. Besides, the hypothesis, even if accepted as a working model, would not directly help in explaining why a single set for a category containing $3N$ members could operate at maximum efficiency, but not a double set for two classes each of which consists of only N members. It would appear that only an adequate model of concept formation could help in answering the question raised. Yet, even if one were available, it would beg rather than answer the question, for concept formation itself occurs only when an appropriate set is in operation. The problem still remains: how are the boundaries determined within which the selective effects of sets operate?

The model presented as being in accordance with the empirical findings of this study (i.e. a model of general inhibition of perceiving all that is irrelevant) is *prima facie* unattractive, for it is uneconomical and contrary to well-established ways of thinking. In fact, it is not inconsistent with previous psychological theory. It makes sense in the analysis of behaviour in certain neuro-psychiatric diseases, and it accords with some recent developments in neuro-physiology, exemplified in the auditory modality by the findings of Hernandez-Péon & Scherrer (1955). Very probably, the hypothesis here advanced will be tenable only in combination with some complementary theory of direct facilitation. It is presented in the most tentative manner; more experimental work along lines similar to those of this study, and others converging from different directions, is needed before this point of view can be stated with greater confidence, or rejected.

REFERENCES

ALLPORT, F. H. (1955). *Theories of Perception and the Concept of Structure.* New York: Wiley.

BRUNER, J. S. & GOODMAN, C. D. (1947). Value and need as organizing factors in perception. *J. Abnorm. (Soc.) Psychol.* **42**, 33–44.

BRUNER, J. S. & POSTMAN, L. (1949). On the perception of incongruity: a paradigm. *J. Personality*, **18**, 206–23.

FREEMAN, J. T. & ENGLER, J. (1955). Perceptual recognition thresholds as a function of multiple and single sets and frequency of usage of the stimulus material. *Percept. Mot. Skills.* **5**, 149–54.

HEBB, D. O. (1949). *The Organization of Behavior.* New York: Wiley.

HERNANDEZ-PÉON, R. & SCHERRER, H. (1955). Habituation to acoustic stimuli in cochlear nucleus. *Fed. Proc.* **14**, 71.

HOWES, D. H. & SOLOMON, R. E. (1951). Visual duration threshold as a function of word probability. *J. Exp. Psychol.* **41**, 401–10.

KOFFKA, K. (1935). *Principles of Gestalt Psychology.* London: Kegan Paul.

KÖHLER, W. (1929). *Gestalt Psychology.* New York: Liveright.

Levine, R. I., Chein, I. & Murphy, G. (1942). The relation of the intensity of a need to the amount of perceptual distortion. *J. Psychol.* **13**, 283–93.

McGinnies, E. (1949). Emotionality and perceptual defense. *Psychol. Rev.* **56**, 244–51.

Miller, G. A., Heise, G. A. & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials, *J. Exp. Psychol.* **41**, 322–35.

Postman, L. & Bruner, J. S. (1949). Multiplicity of set as a determinant of perceptual behavior. *J. Exp. Psychol.* **39**, 369–77.

Postman, L., Bruner, J. S. & McGinnies, E. (1949). Personal values as selective factors in perception, *J. Abnorm. (Soc.) Psychol.* **42**, 142–54.

Postman, L. & Crutchfield, R. S. (1952). The interaction of need, set, and stimulus-structure in a cognitive task. *Amer. J. Psychol.* **65**, 196–217.

Postman, L. & Leytham, G. (1951). Perceptual Selectivity and ambivalence of stimuli. *J. Personality*, **19**, 390–405.

Thorndike, E. L. & Lorge, I. (1944). *The Teacher's Word Book of* 30,000 *Words.* New York: Columbia University Press.

Wilcoxon, F. (1949). *Some Rapid Approximate Statistical Procedures.* New York: Amer. Cyanamid Co.

(*Manuscript received* 27 *February* 1957)

TACHISTOSCOPIC EXPOSURE TIME AND SPATIAL PROXIMITY IN THE ORGANIZATION OF VISUAL PERCEPTION

By J. W. KASWAN

Psychology Department, Yale University

This study investigated the temporal sequence of spatial organization, using varying proximity of geometric designs (e.g. (h) .. .. .. and (d) .. .. ..) as the criteria for organization. In addition, the study explored the possible impairment of such space-time relations in schizophrenics. Perception of groupings evolved in a predictable sequence of differentiation, correct groupings (e.g. pairs of dots) emerging more quickly the better the spatial differentiation. The results were similar for twenty-four medical-surgical and twenty-four schizophrenic hospital patients, except that the latter required more time to see correct groupings in most of the designs.

I. Introduction

It is generally recognized that visual perception occurs in time, and the literature of recent years abounds in studies on the relation of temporal factors to intensity of stimulation (Hunter & Sigler, 1940), the span of 'discrimination' (Gilliland, Hofeld & Eckstrand, 1946), and other phenomena. The present study concerns two related problems. One is the relation of exposure time to the phenomenology of spatial organization, as characterized by simple geometric patterns varying in the proximity of grouped units. The other is the possible impairment of the time sequence of perceptual organization in schizophrenics.

As regards the first problem, it is generally agreed that some of the grouping factors emphasized by Gestalt psychologists play an important role in perception. Some of the designs originally used by Wertheimer (1950, p. 88) to demonstrate the role of grouping in perception were adapted by L. Bender for inclusion in a test designed to measure perceptual-motor integration. She found that when these designs were presented to normal subjects at fast tachistoscopic exposures the resulting percepts were similar to those observed in small children and some schizophrenics, for whom exposure time had not been reduced. Her work led her to suggest that a time factor may be involved in the organization of perception and that this organization may follow some orderly sequence of differentiation in which a dextrad horizontal ordering of events is an early stage (Bender, 1938, pp. 51–52). Since visual perception usually involves complex groupings of details, determination of time parameters and their relation to spatial characteristics is important not only for a better understanding of perception, but perhaps also for the elucidation of some aspects of behaviour in psychopathology. Thus, there is evidence that schizophrenics show much longer reaction times than normals in a variety of tasks and a basic disturbance of adaptive capacity has been postulated to account for this deficit (King, 1954, p. 155). Since perceptual activity is the first link in the organism's relation to the environment, an increase in the time required for perceptual organization may well be part of the reason for the general sluggishness of reactive behaviour characteristic of such patients.

With spatial proximity as the criterion for perceptual organization, it was assumed that perception of groupings in time would follow a process of differentiation. Specifically, it

was proposed that: (*a*) the clearer the grouping, the less time should be required to perceive it; (*b*) schizophrenics should require a longer time to perceive correct groupings; and, (*c*) the process of differentiation should generally proceed in the following sequence: indistinct blur (e.g. bunch of dots)—dextrad horizontal line of objects—incorrect or unclear groupings—correct groupings.

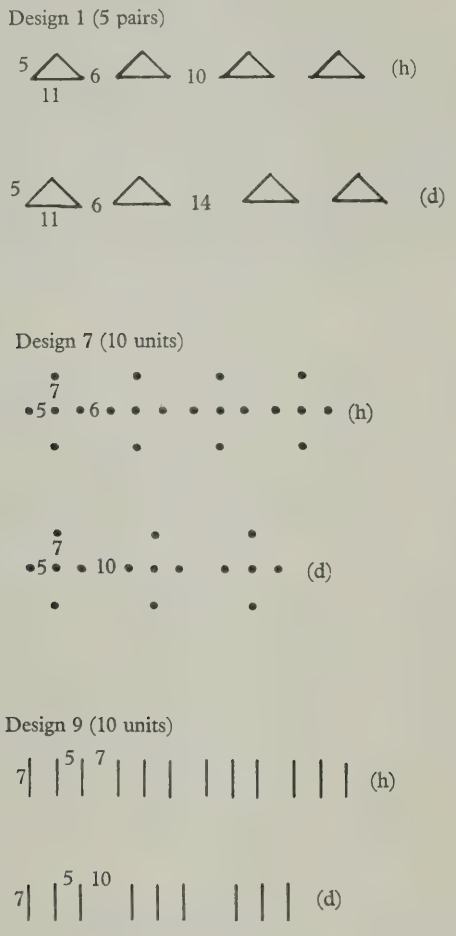


Fig. 1. Examples of designs. Numbers refer to measurements in millimetres of actual size.

II. Methods and procedure

(i) *Materials.* There were a total of twenty-one different designs, some of which are shown in Fig. 1.* All designs were drawn through the middle of 9×12 in. white cards, across the field of vision in the tachistoscope, while a 4 in. margin at the right and a $1\frac{1}{2}$ in. margin at the left fitted into the frame of the apparatus. There was a practice series of nine designs and the main experimental series of twelve basic designs. Ten of the twelve designs consisted of two variants differing in the proximity of grouped units, making a total of twenty-two experimental designs. Where the groupings were closer together, the designs

* To save printing costs samples of all the designs used and the predicted A, B and C responses for each have been deposited with the Technical and Documentary Unit, Department of Scientific and Industrial Research. The materials may also be obtained directly from the author.

were deemed more homogeneous (h); where further apart, they were better differentiated (d). Two were control designs containing simple line drawings without intended grouping and were included, in part, to check on possible guessing tendencies. Of the nine practice designs four were (h)-type drawings, three were (d)-type drawings, and two were control designs containing no grouping. The designs were selected on the basis of the results of a pilot experiment with about thirty drawings and were those the subjects could most clearly describe and reproduce.

To determine whether the intended groupings were easily and clearly visible in both (h) and (d) versions when sufficient time was allowed for their perception, the experimental designs were presented to two groups of 15 adult subjects who were asked individually to describe each drawing within 5 sec. and then rate them 'very clear', 'fairly clearly', or 'not clear'. Each group was given a random arrangement of five (h) and five (d) designs. The results indicate that 97·2% perceived (h) designs as 'very clear' and 95·5% saw (d) designs as 'very clear', thus demonstrating the unambiguous nature of the designs.

(ii) *Subjects.* There were forty-eight male subjects, all Veterans Administration hospital patients, divided into two groups of twenty-four each, classified as Medical and Surgical (MS) and Schizophrenics (Sch). Selection was by chart diagnosis and no subjects with known brain disturbance or visual difficulties were selected. The schizophrenic subjects were all described as 'chronic' and the ward psychiatrist believed none to be in a state of complete remission. Subjects who had undergone electric shock therapy within 3 months prior to the experiment were excluded. The two groups did not differ significantly in age, education, or intelligence as measured by the Wechsler–Bellevue Block Design and Vocabulary subtests.

(iii) *Apparatus.* A Gerbrands mirror tachistoscope was used on which exposure time could be set from 0·01 to 1 sec., at 0·01 sec. intervals. (This instrument provides a constant level of illumination.) For periods exceeding one second a stop-watch was used. A memory drum was adapted as a recording device on which subjects could draw what they saw without seeing their drawing. The experimenter turned a knob which moved the position of the continuous strip after each drawing.

All subjects' eyesight was tested with a standard Snellen chart and only those with monocular vision of 15/30 or better continued in the experiment.

(iv) *Procedure.* After the test of visual acuity, the practice designs were presented in the tachistoscope with the following instructions. 'In a moment I will show you some drawings of lines or dots in different arrangements. As soon as each drawing appears tell me what it looks like as carefully as you can and at the same time draw it right here. As you won't see what you are drawing, just try to draw so that I can get a good idea of what you saw. Any questions? Let's try it.' Subjects who hesitated were urged to give their first impression; any response delayed more than 5 sec. after presentation was scored as failed. The nine cards of the practice series were all presented first at 0·5 sec. exposure and then again, in the same sequence, for 5 sec. The purpose of the practice series was not only to acquaint subjects with the task to be performed, but also to induce a set to see groupings in the designs. To increase the probability that all subjects had a substantial set to see groupings, only those remained in the experiment who could accurately perceive the correct grouping in the last three (h)-type designs of the practice series.

The experimental series was introduced after the practice series without further instructions. As can be seen from the description of the materials, every design had two variants, differing in the distance between grouped units. Each variant was presented separately, so as to get relatively independent results for the two, and it was therefore necessary to divide the twenty-two designs into two groups, series A and B, each containing 5 (h), 5 (d), and one control design. Designs 1h–5h, 6d–10d and 11 were in series A, the remainder in series B. All subjects were given both series at a 3- to 7-day interval between presentations. Subjects were told nothing about the second presentation at the time of the first. Half the subjects in each group were given series A first, the other half were given series B first.

Six exposure times were selected for each design on the basis of the responses of normal subjects in the pilot experiment, divided into six percentage groups. This was done to minimize the possible effect of rigidity on the part of schizophrenic subjects which had been demonstrated in the pilot experiment. Exposure times varied from design to design but were the same for (h) and (d) variants of the same design. All cards were first presented to each subject at the fastest exposure time (0·01 sec.). Then they were rearranged in a different order and presented at the next, longer, exposure time, and so on, to the sixth exposure time. Exposure times, in seconds, for the designs shown in Fig. 1 were as follows: design 1, 0·01, 0·04, 0·10, 0·25, 0·75, 5 and 10; for design 7, 0·01, 0·06, 0·15, 0·25, 0·75, 5 and 25; for design 9, 0·01, 0·06, 0·10, 0·50, 0·90, 5 and 20. The order of exposures followed an 11×11 Fisher Latin square. Serial position was also different in the two series shown to each subject in order to avoid exposure of

(d) and (h) variants in the same sequence as the previously shown (h) and (d) variants, but there may have been occasional coincidences. It was hoped that this procedure would minimize any possible learning due to previous experience with the designs.

(v) *Scoring.* General criteria, applicable to all designs, were as follows. 'A' scores: (*a*) no grouping at all was perceived; (*b*) the response was entirely unrelated to the stimulus. 'B' scores: (*a*) perception of some vague groupings; (*b*) groupings that were partly incorrect. 'C' scores: accurate description of the shape and contents of the design. These were the only correct responses and, unless otherwise noted, all statistics are based on the first of three successive 'C' responses to every design. In addition, there were specific scoring criteria for each design. For instance, design 9 (see Fig. 1) has the following criteria: 'A'—line of vertical lines, 'B'—groups of two or four lines; and 'C'—groups of three vertical lines. Responses were recorded verbatim and scored after each individual session.

III. Results

In the practice series there was no consistent difference between the groups for perception of the correct grouping on the different designs at the 0·5 sec. exposure. At the 5 sec. exposure, all subjects who continued in the experiment gave correct responses, with one exception in each group for one of the early practice designs. Four schizophrenics were eliminated from the experiment because they could not meet the minimum requirement, which was to give an adequate verbal description of the correct grouping of the last three (h) designs in the practice series, and it seems probable that all subjects had a set to see distinct groupings in the main experiment.

Table 1 summarizes the significance of the mean difference between the ordinal position of exposure at which correct perception of (h) and (d) designs occurred. It was assumed that it would take longer to perceive (h) designs (higher ordinal position) and the results clearly support this hypothesis, as the mean difference from zero is significant in the

Table 1. *The significance of the mean difference between the ordinal position of the correct perception of groupings in normals and schizophrenics*

	Normals		Schizophrenics	
Designs	Mean diff.	*t*	Mean diff.	*t*
1h–1d	2·33	6·51***	2·08	7·25***
2h–2d	1·33	5·81***	1·46	3·47**
3h–3d	2·01	5·30***	2·17	5·61***
4h–4d	1·62	3·63**	2·75	5·86***
5h–5d	2·62	6·28***	3·08	7·32***
6h–6d	1·08	2·91**	1·67	4·41***
7h–7d	1·50	3·77***	1·92	4·27***
8h–8d	0·33	0·82	1·67	3·74***
9h–9d	2·04	5·70***	2·50	6·13***
10h–10d	0·67	1·77*	0·54	1·64

* $P > 0{\cdot}05$. ** $P > 0{\cdot}01$. *** $P > 0{\cdot}001$.

expected direction in eighteen out of twenty comparisons. Fig. 2 shows the occurrence of 'C' responses in actual time units for the designs given in Fig. 1. While the gradient of the curves varies with the design, the difference between (h) and (d) variants is consistently maintained, most curves approximating the ogive form characteristic of psychophysical functions. Table 2 shows that for thirteen out of twenty Kolmogorov–Smirnov comparisons (Goodman, 1954) schizophrenics required a significantly longer exposure time to see the correct groupings. All differences were in the expected direction. Almost

all 'A' and 'B' responses corresponded to the *a priori* criteria established for each design. Only 3·24% of these 'primitive' structure responses seemed idiosyncratic and most of these were given by three schizophrenic patients.

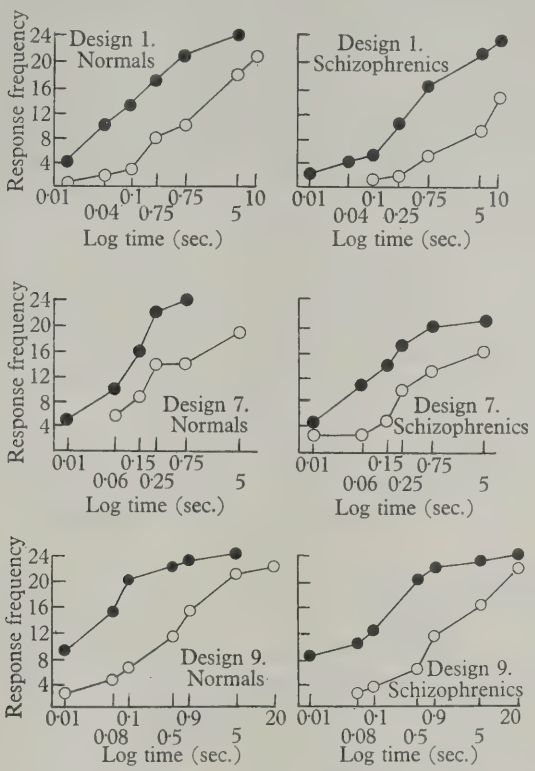


Fig. 2. Cumulative frequency distribution of correct (C) responses for three representative designs. *Key*: blacked circles represent responses to (d) variants, open circles refer to (h) variants of each design.

Table 2. *Kolmogorov–Smirnov test for the significance of the difference between ordinal position of correct perception of groupings in schizophrenics and normals*

Design	$D \times N$	P	Design	$D \times N$	P
1h	9·12	0·03	6h	9·02	0·03
1d	8·16	0·08	6d	5·04	0·30
2h	10·08	0·02	7h	3·84	0·40
2d	9·12	0·03	7d	5·04	0·30
3h	11·04	0·01	8h	10·10	0·02
3d	12·24	0·01	8d	4·01	0·40
4h	9·12	0·03	9h	5·04	0·30
4d	9·01	0·03	9d	8·16	0·08
5h	13·92	0·001	10h	12·96	0·01
5d	9·95	0·03	10d	10·08	0·02

Initial responses were usually 'A' responses, followed by 'B' and 'C' responses. This is suggested by Fig. 3 which indicates a steady decrease in the amount of 'A' responses from the first to the last exposure, with the reverse trend in the case of 'C' responses.

As expected, 'B' responses tend to be perceived later than 'A' responses. All the differences are statistically significant. Comparison of the 'A' and 'B' responses yield a χ^2 of 20·51 for the medical-surgical subjects and of 12·90 for the schizophrenics, significant

at beyond the 0·01 and the 0·05 level respectively. The difference of the trend in 'C' responses compared with 'A' and 'B' responses is too obvious to require computation.

Since the sequence of exposures was from fast to slow, it seems reasonable to suppose that the development of groupings as here described follows the predicted process of differentiation from homogeneity. Since exposures were continued until there were three successive 'C' responses, 'A' or 'B' responses occasionally intervened between the first and successive 'C' responses. The results show that 3·3% of medical-surgical and 11·8% of schizophrenic responses were of this type. The relatively larger number of such responses by schizophrenics support frequent findings that these subjects tend to be more erratic than non-psychiatric subjects.

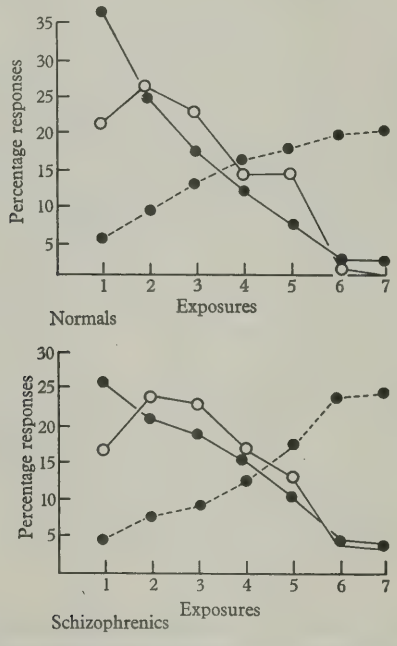


Fig. 3. Percentage of A, B and C responses at each exposure. *Key*: blacked circles with continuous line responses represent A responses. Open circles refer to B responses and blacked circles with dashes refer to C responses.

The two control designs, one each in series A and B, which contained no grouping, were correctly seen at the first exposure by 94% of subjects, indicating that the time factor in the present investigation refers to groupings and not to the perception of figures as such.

Correct perception of corresponding (h) and (d) variants is highly and significantly correlated. The Pearson r for medical-surgical subjects is 0·804 and for schizophrenics 0·982. This seems to indicate that the designs used were fairly constant stimulus patterns, changes in proximity of (d) variants affecting performance only in relation to (h) variants.

IV. Discussion

The first problem of this study concerns the relation of exposure time to the phenomenology of spatial organization. The results clearly suggest that the perception of spatial groupings is predictably different at various time exposures. This is demonstrated in two ways. On the one hand it took significantly longer to identify correct groupings in (h) than (d) designs, supporting our expectation that decreased proximity between groupings, favouring horizontal line arrangements, would impede perception of groupings under reduced time exposure relative to (d) variants. On the other hand, it was shown that at short exposure times subjects perceived 'a bunch of lines', or the dextrad horizontal ordering of elements, such as 'bunch of vertical lines', ('A' responses) where vertical lines in groups of three (design 9) would have been the correct percept. At longer exposures in the tachistoscope some grouping was perceived, such as 'dots in groups of threes with some scattered on top and on the bottom' (design 7), and finally, at even longer exposures, most subjects perceived the correct grouping. Thus the results indicate not only that perception of groupings develops in time, but, what is more important, they also provide some clues as to *how* this development occurs. These findings gain added validity from the fact that the differences between (h) and (d) are consistent for two groups of subjects who differ in the total amount of time each required for the perception of correct groupings.

The second problem of this study concerns the schizophrenics' impairment of space-time relations in perception. The longer times required by these subjects to see the correct groupings support the assumption that their long response latencies extend to this very basic level of the organism's translation of the environment. Acquisition of knowledge about the environment within a limited range of inter-individual constancy—perceiving what most other people see—must be at least partly what clinicians call 'sense of reality'.

The time relations explored in this study should be clearly distinguished from those often described by psychoanalysts when they discuss time functions in psychopathology. They speak of the schizophrenic's inability to delay expression of needs (Bonaparte, 1940). The present study deals with time disturbance in the *opposite* direction—the slowness with which schizophrenics perceive groupings. The former involves primarily a defect in the inhibition of need expression, the latter, sluggishness in the pacing of reactions to the environment. While this illustrates that different time parameters may play very different roles in behaviour, only further experimental work—now in progress—can hope to clarify their relationship, as well as their connexion with other cognitive functions, such as learning.

This report is based on a dissertation submitted to the Graduate School of Washington University, St Louis, Missouri, in partial fulfilment of the requirements for the Ph.D. The writer is greatly indebted to Dr Saul Rosenzweig, under whose guidance this project was undertaken. Grateful acknowledgement is also made to Drs Marion E. Bunch, R. M. Collier and Frederick H. Kanfer for their valuable suggestions and criticisms. The co-operation of Veterans Administration hospitals and clinics in St Louis and Jefferson Barracks, Missouri, and Little Rock, Arkansas, is gratefully acknowledged.

References

Bender, L. (1938). *A Visual Motor Gestalt Test and its Clinical Use.* New York: Amer. Orthopsychiat. Association.

Bonaparte, M. (1940). Time and the unconscious. *Int. J. Psycho-Anal.* **21**, 427–68.

Gilliland, A. R., Hofeld, J. & Eckstrand, G. (1946). Studies in time perception. *Psychol. Bull.* **43**, 162–76.

Goodman, L. A. (1954). Kolmogorov–Smirnov tests for psychological research. *Psychol. Bull.* **51**, 160–8.

Hunter, W. S. & Sigler, N. (1940). The span of visual discrimination as a function of time and intensity of stimulation. *J. Exp. Psychol.* **26**, 160–79.

King, N. E. (1954). *Psychomotor Aspects of Mental Disease.* Cambridge, Mass.

Wertheimer, M. (1950). Laws of organization in perceptual forms. In *Sourcebook of Gestalt Psychology*, E. D. Ellis (ed.). New York.

(*Manuscript received* 28 *March* 1957)

PSYCHOMOTOR CONTROL IN STAMMERING AND NORMAL SUBJECTS: AN EXPERIMENTAL STUDY

By GEORGE SETH

Psychology Department, The Queen's University, Belfast

Voluntary control of the regularity of a motor performance (finger-tapping) which does not involve speech or the speech mechanism was investigated in fifteen stammering and fifteen non-stammering subjects. The results indicate: (i) that in right-hand finger tapping the stammerers are markedly inferior to normal controls in their ability to maintain voluntarily the rhythm or regularity of the performance; (ii) that this group of ostensibly right-handed stammerers showed a marked inferiority in right-handed, as against left-handed, performance; (iii) that the disorganization of psychomotor performance is not confined to the speech-function, and may appear in situations where communication by speech is not involved, and in which the influence of personal-social factors is reduced to a minimum.

I. Introduction

It has long been argued by neurological and neuropsychological investigators (Travis, 1931) that the emotional and situational factors that are so obviously influential in stammering speech are secondary and that 'in the great majority of cases the basic cause of stuttering (stammering) is to be sought in some significant variation in the functional integration of the central nervous system'. This view of the stammering disorder, in terms of a physiological rather than a psychological mechanism, derives from a theory of cerebral dominance that has been heavily criticized (Burt, 1937), but it has continued to influence research and treatment up to the present time. In particular, it implies, first, that disturbance in the control of the speech-mechanism should be apparent not only in the speech-situation, and secondly, that in a considerable number of cases of stammering it should be possible to demonstrate inadequate development of the normal preferential sidedness (handedness); or interference, in the 'natural' left-hander, with its establishment. The experimental studies of Blackburn (1931) and Seth (1934) provided evidence for the first of these two hypotheses. These reports showed quite unequivocally that, even in situations in which speech is not required and in which the emotional temperature has been kept as low as possible, control of the mechanism of speech and especially of that of respiration is significantly more erratic in stammering subjects than in normal-speaking controls. Blackburn's results indicated that voluntary control of movements of the diaphragm, lips, jaws and tongue were all defective in the stammerer, and further evidence (Seth, 1934), while showing no statistically significant differences for movements of the peripheral musculatures, was even more emphatic in respect of the disorganization of voluntarily sustained respiratory movement.

In addition to his experiments on control of the speech-mechanism, Blackburn carried out a subsidiary inquiry designed to compare the performance of stammerers and normal speakers in voluntarily maintaining regular movements of structures other than those of speech. His results revealed no significant difference between the two groups of subjects. Indeed, the most striking feature of Blackburn's tables is that all his subjects, stammering and normal alike, behaved in his tests of tapping-rate virtually like human metronomes,

showing in the great majority of the records zero standard deviations. It is hardly credible that human subjects could have maintained such a standard of regularity, and it may be questioned whether his method of measuring his kymographic records (which he does not specify) was not too crude to reveal the variations within and between the individual tracings. Since there are some apparent errors of computation in his presented data (Blackburn, 1931, table IV), it is not clear in what sense they should be taken. But it is certainly noteworthy that a reworking of his figures suggests a conclusion, although the differences are insufficient to substantiate it, that runs counter to the laterality theory that had been popularized by the Iowa laboratory where his work was done. So far as variation (i.e. erratic control) in tapping rate appeared at all in Blackburn's group, it was greater for the right hand than for the left in his normal subjects, and for the left hand than for the right in the stammering series.

II. The present experiments

Despite Blackburn's apparently negative results for tapping performance, there seemed to be some good reasons for investigating the matter more thoroughly than he had done. These were: (i) the intrinsic interest of the problem of 'sidedness', and the difficulty of arriving at a satisfactory test for the basic preference of the individual subject; (ii) the hypothesized relation between cerebral dominance, 'handedness', and the stammering disorder; and (iii) the suggestion, even in Blackburn's unsatisfactory data, of a difference between the stammerers and the normal speakers in respect of left- and right-hand control.

Consequently, in the present experiment the tapping test was run for both hands with two groups of subjects, consisting of fifteen ostensibly right-handed stammerers and fifteen right-handed normal controls. The subjects were children between the ages of 11 and 15; in each group there were two girls, and the groups were approximately matched for chronological age. The problem was to determine whether there are significant qualitative and quantitative differences in the unilateral motor control of stammerers and non-stammerers.

Kymographic records were obtained of three trials, each of approximately 15 sec. for each subject for each hand, giving an over-all recording time of 45 sec. for each hand. Subjects were required by the instructions to tap 'fairly quickly' on a tambour with a moderately thick rubber diaphragm on which was mounted a metal disk. During all trials the arm was unsupported, and right- and left-hand trials were taken alternately. The subjects were instructed to maintain, so far as possible, an even rate of tapping throughout the performance, but speed was not emphasized. The experimenter demonstrated what was required, and 15 sec. of unrecorded practice were allowed at the outset. The apparatus was concealed so that the subject could not regulate his performance by watching his own record.

Evaluation of the records

(1) Evaluation of the records required a count of the tapping movements recorded within a given time interval, fractional parts of a movement being assessed with reasonable accuracy by means of an opisometer. The section of the record under scrutiny was marked off by two vertical strands of fuse-wire stretched across a rectangular frame which could be moved along the time line like the cursor of a slide-rule. (2) In each 15 sec. record the first 3 sec. were ignored so that the subject might have had time to 'get into his stride' and establish a natural rhythm of performance. (3) The chief quantitative variable in which we are interested being *regularity* of performance, an appropriate sample of the subject's product was obtained by counting the movements for every third second of the record, time being recorded throughout in half-seconds.

The data for the individual subjects are set out in Table 1. In summary form, the relevant statistics from the two groups are as follows:

	Mean co-efficient of variation	
	Right hand	Left hand
Stammering group	9·39	6·17
Non-stammering group	4·87	5·16

Table 1. *Tapping records*

Subject ...	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Stammering group: (*a*) right hand															
M.	2·74	2·96	3·63	3·10	3·77	3·66	2·21	2·70	3·27	2·92	3·26	2·84	2·89	3·59	3·04
S.D.	0·37	0·42	0·15	0·28	0·33	0·50	0·15	0·20	0·17	0·27	0·41	0·28	0·17	0·39	0·27
V.	13·76	14·18	4·13	9·03	8·75	13·66	6·79	7·41	5·20	9·24	12·81	9·77	6·38	10·86	8·88
Stammering group: (*b*) left hand															
M.	2·54	2·85	3·39	2·92	3·31	3·10	2·24	2·14	3·14	2·92	3·14	2·72	2·66	3·14	2·93
S.D.	0·19	0·24	0·14	0·16	0·40	0·14	0·11	0·18	0·18	0·15	0·22	0·23	0·15	0·19	0·18
V.	7·31	8·42	4·13	5·48	12·08	4·52	4·91	8·41	5·73	5·13	7·16	8·34	5·94	6·05	6·20
Non-stammering group: (*a*) right hand															
M.	2·75	2·17	2·80	2·07	2·67	2·31	2·24	2·94	2·79	2·05	2·68	3·16	2·82	2·03	2·58
S.D.	0·24	0·14	0·10	0·04	0·14	0·12	0·16	0·12	0·11	0·09	0·11	0·15	0·11	0·08	0·14
V.	8·62	6·45	3·60	1·93	5·24	5·19	7·14	4·08	3·94	4·15	4·22	4·86	4·06	3·95	5·61
Non-stammering group: (*b*) left hand															
M.	2·45	2·43	2·73	2·11	1·90	2·97	2·01	3·16	2·25	2·24	2·21	2·78	2·53	1·97	2·46
S.D.	0·18	0·22	0·10	0·06	0·10	0·30	0·06	0·15	0·12	0·12	0·09	0·14	0·10	0·08	0·09
V.	7·22	9·05	3·62	2·84	5·26	10·10	2·99	4·74	5·33	5·35	4·08	5·02	4·13	4·06	3·66

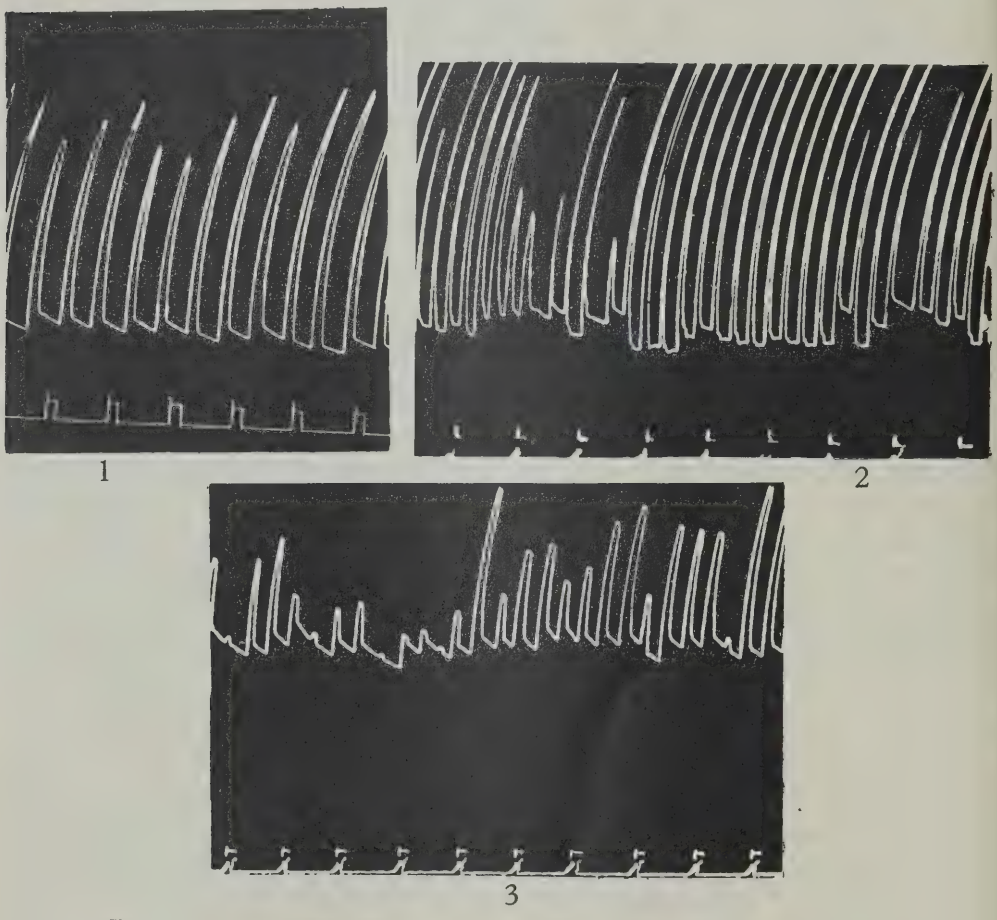
M. = average movements per second; V. = coefficient of variation (100 S.D./average).

III. Discussion

Earlier experiments have shown that stammering subjects display a higher degree of variability in control of different parts of the speech musculature than do normal speakers. But some studies have tended to suggest (Seth, 1943) that this inferiority might even there be confined to the mechanism of respiration. The present result, however, makes it clear that, in our experimental group, the erratic control is not confined to the speech situation nor even to the speech mechanism, but extends to other forms of psychomotor performance as well. In particular, the following points should be noted. First, in the stammering group the coefficient of variation (i.e the inferiority in rhythmical performance) is very much greater for the apparently preferred right hand and arm than for the left; in the normal group, while there is no marked difference, the superiority lies in the preferred right hand. Using the Wilcoxon Matched Pairs Test, $p < 0{\cdot}01$ and $p > 0{\cdot}05$ respectively. Secondly, the coefficient of variation is greater for both hands in the stammering group, but it is very much greater when the right hands in particular are compared—9·39 as against 4·87. The Median Test gives: (*a*) right-hand, $p = 0{\cdot}0013$; (*b*) left-hand, $p = 0{\cdot}048$. Thirdly, examination of the individual results serves to emphasize both of these points. In only five cases, using the right hand, do the individual stammerers show a coefficient of variation lower than that of the most irregular of the control subjects, whereas for the left hand, in all but one of the stammering subjects, the records fall within the range of the normal group.

These results are clearly consistent with a neurophysiological theory of the stammering disorder on Orton–Travis lines. The disorganization of speech that we call stammering is certainly highly susceptible to changing features of the social situation in which speech is demanded; but then so also are the speech and other everyday motor performances of the patient suffering from cerebral lesion in the dominant hemisphere. It may be, therefore, that in the one condition as in the other the disorganization of nervous function must, in a proportion of the cases at least, be regarded as fundamental. Our results suggest, moreover, that stammering is not to be considered merely as a disorder of speech

but that, at least by early adolescence, the condition is demonstrable in some other aspects of psychomotor function, where it is ordinarily unsuspected. In so far as this implies a measure of consistency among the expressive acts of one stammering subject, it merits consideration in the light of the Allport–Vernon studies of expressive movement (Allport & Vernon, 1933). Their general conclusion is, indeed, so relevant to these results that it may be quoted directly. They say: 'Motor acts are not so specific as to be meaningless, and being organized they must reflect to a large degree the organization of the total brain-field. There are degrees of unity in movement, just as there are degrees of unity in



Figs. 1, 2 and 3. Tapping records. 1, normal speaker; 2 and 3, stammering speakers.

mental life and in personality. It is surely not unreasonable to assume that in so far as personality is organized, expressive movement is harmonious and self-consistent, and in so far as personality is unintegrated, expressive movement is self-contradictory.' Perhaps, however, the most important implication is this: that, while there is general agreement that the difficulty in the stammering disorder is central or cortical, our current attitudes, as they find expression in theory and in therapy alike, too often involve yet another instance of that 'neglect of the effectors' for which psychology in general has

been rightly criticized, and more energetic pursuit of inquiry into a variety of aspects of psychomotor function in a wider range of groups, especially of so-called 'functional' or organic disturbance, might be rewarding.

Some typical records are reproduced which serve to illustrate the foregoing points. These may be compared with those presented for respiration, etc., in the earlier studies (Blackburn, 1931; Seth, 1934) to which reference has been made. They provide concrete and striking qualitative evidence for the points that are summarized and generalized in the tabulated data. Fig. 1 is taken from the middle of the third record of tapping with the index finger of the right hand, produced by the most regular of the non-stammering subjects (no. 14, S.D. 0·08). The quantitative and qualitative uniformity of the record is particularly striking, and the regularity in the height of the separate peaks of the tracing, which answers to the force with which each of the separate taps was delivered, is noteworthy. A corresponding record by a stammering subject (no. 1, S.D. 0·37; third trial with the right hand) is presented in Fig. 2, the inferiority of the control being apparent in the general arhythmical 'jerkiness' of the tapping and the wide range of variation in the force of the percussion. Qualitative irregularity is even more obvious in Fig. 3, in which is presented part of the performance of another stammering subject (no. 14, third trial with the left hand) whose quantitative record is not by any means unique (S.D. 0·19). But the standard deviation, depending upon count alone, perforce must disregard, and in some cases obscure, the haphazard manner in which the tracing (i.e. the forcefulness of the tapping) fluctuates from moment to moment, with no appearance of periodicity.

REFERENCES

ALLPORT, G. W. & VERNON, P. E. (1933). *Studies in Expressive Movement.* New York: Macmillan.

BLACKBURN, B. (1931). Voluntary movements of the organs of speech in stutterers and non-stutterers. *Psychol. Monogr.* **41**, 4.

BURT, C. (1937). *The Backward Child.* London: University of London Press.

SETH, G. (1934). An experimental study of the control of the mechanism of speech in stuttering subjects. *Brit. J. Psychol.* **24**, 375–88.

TRAVIS, L. E. (1931). *Speech Pathology.* New York: Appleton and Co.

(*Manuscript received* 8 *April* 1957)

COGNITIVE AND SOCIAL CHANGES IN THE FEEBLEMINDED —THREE FURTHER STUDIES

By A. D. B. CLARKE, A. M. CLARKE and S. REIMAN

Psychology Department, The Manor, Epsom, Surrey

Earlier papers have described researches showing that I.Q. increments in adolescent and young adult persons certified as feebleminded occur not infrequently, cannot be explained as artefacts, and appear to be linked with early very adverse environmental experiences. The present three studies indicate that I.Q. increments can be large over long periods of time, are correlated with changes in social adjustment, but cannot be shown to be induced by special environmental stimulation; in the present age range and type of subject, they tend to occur in any reasonable environment. Results imply that a necessarily poor outcome cannot be predicted for children with I.Q.'s in the 50's, 60's or 70's if they also come from very adverse conditions. Test reliability, statistical regression, and a review of some relevant literature are discussed; theoretical problems concerning intellectual development are presented; and attention is directed to individual differences in susceptibility to environmental damage and to the whole problem of human resilience.

I. Introduction

In earlier papers (Clarke & Clarke, 1953, 1954*b*, 1955) the general problem of I.Q. inconstancy was considered and, specifically, the variability of the I.Q. among adolescents and young adults certified as feebleminded was discussed. A two-year study indicated that I.Q. increments were of frequent occurrence, particularly in cases characterized by extreme deprivation in early life. In a population drawn, on the whole, from very adverse socio-economic conditions, those from the worst had made an average increment of 9·7 points, while a residual group from somewhat less unfavourable conditions made an average increase of 4·1 points. This latter was similar in amount, though not necessarily of the same nature, as the gains arising from the maximal effects of test practice and errors of underestimation in a control group retested after a short time interval. Cognitive changes in the certified feebleminded differed from those occurring in normals in that: (1) they tended to be mainly in the direction of improvement; (2) in relation to time interval between test and retest, they tended to be rather large; and (3) they took place at ages when intellectual growth is commonly assumed to have ceased. The general picture which emerged from this and other studies was that many high-grade defectives showed temporary arrest of intellectual growth, while later reverting towards (and sometimes achieving) normality.

When the earlier work (which throughout the present paper will be termed the *original research*) was completed, it was decided to plan longer-term investigations which might throw further light on the determinants and limits of these increments, and which might have further implications for our understanding of mental deficiency. Objective psychometric assessments have been carried out at this hospital since 1949, and it was considered that by 1955–6 sufficient data would be available to make such longer-term research possible. In all, three new studies are reported here.

II. First Study

Cognitive changes

The first experiment was designed to follow-up those members of the *original research* who were still available. The original subjects had, in 1952, been allotted either to a group from very bad homes (based on the presence of two or more of twelve criteria* in the case history) or to a residual group from less adverse circumstances; this rating had been carried out independently by an external investigator who knew neither the patients nor their test scores. It seemed clear that the 28 available at the end of 1955 (from an original 59) were, in the main, the less satisfactory members of the sample, since a large proportion

Table 1. *Test-retest data, Wechsler, Form I,† over 6 years*

	1949, Full Scale I.Q.	1952, Full Scale I.Q.	1955, Full Scale I.Q.	Diff. '49/'52	Diff. '52/'55	Diff. '49/'55
			(*a*) Group from very bad homes			
Mean	59·6	70·7	75·8	11·1	5·1	16·2
	(66·4)	(76·1)		(9·7)		
S.D.	9·7	11·4	10·3	4·2	5·0	6·1
	(14·1)	(13·4)		(6·3)		
N	9	9	9	9	9	9
	(25)	(25)		(25)		
t (1-tail test)	—	—	—	7·929	3·059	7·969
Significant at	—	—	—	0·1%	1%	0·1%

I.Q. range: 1949, 47–72; 1952, 58–86; 1955, 64–91.
Age at final retest: mean 26·4, S.D. 5·5 years.
Period between first and final test: mean 70·9 months, S.D. 2·5 months.
Range of I.Q. changes: +3 to +22 points. (10 points and above: 8/9 = 89%; 15 points and above: 7/9 = 78%.)

			(*b*) Group from less adverse homes			
Mean	62·3	66·8	72·5	4·5	5·7	10·2
	(66·1)	(70·1)		(4·1)		
S.D.	13·4	13·0	13·3	4·4	5·1	6·6
	(14·1)	(13·1)		(4·9)		
N	19	19	19	19	19	19
	(34)	(34)		(34)		
t (1-tail test)	—	—	—	4·460	4·872	6·737
Significant at	—	—	—	0·1%	0·1%	0·1%

I.Q. range: 1949, 35–87; 1952, 40–89; 1955, 48–100.
Age at final retest: mean 27·9, S.D. 9·5 years.
Period between first and final test: mean 74·2 months, S.D. 5·3 months.
Range of I.Q. changes: −2 to +21 I.Q. points. (10 points and above: 10/19 = 53%; 15 points and above: 5/19 = 26%.)
t (between group from very bad homes and group from less adverse homes) = 2·403 (significant above 2·5% level, 1-tail test).

(54%) of the remainder had by now been discharged either as being 'no longer mentally deficient within the meaning of the Act' or as being sufficiently competent socially to warrant this action, or alternatively, were on distant licence and likely to achieve discharge in the near future. Table 1 (*a*) and (*b*) indicate that the present subjects were, both in 1949 and in 1952, on the average less intelligent than the remainder of the sample

* The main point in each of these criteria was as follows: N.S.P.C.C. intervention, parental attitude antagonistic, no fixed abode, 'Fit Person' Order, home conditions bad, considerable neglect, irregular school attendance due to neglect, home dirty and neglected, gross poverty, crime in parents, rickets, child found begging.

† Fifty patients drawn from the present three studies were, in addition, given the Progressive Matrices test to confirm their present levels. These correlated to the extent of 0·81, with a mean difference of 2·67 points between them in favour of the Wechsler.

now no longer available, and this fact adds support to the belief that these were indeed the less satisfactory members of the original sample. Thus, the results of a follow-up of those remaining would be expected to give a minimal estimate of the I.Q. change over a period of 6 years for patients of this age, type and social background. Table 1(*a*) and (*b*) show the test and retest data on Wechsler, Form I, for the members of the two groups. The figures in brackets indicate the relevant data for the original groups of which the present subjects formed a part.

It will be noted that the I.Q.'s of those from very bad homes present a substantially different cognitive picture in 1955 from that in 1949, having, on the average, increased by about one standard deviation. The I.Q.'s of those from less adverse home circumstances had changed by about two-thirds of a standard deviation, and it is clear that in their case the change had been slower, smaller in amount, and had been roughly equal in the two equivalent time periods, 1949–52 and 1952–5. Members of the former group, however, had shown the maximum change in the earlier part of this period, thereafter improving less rapidly. This does not conflict with the earlier hypothesis that I.Q. increments seemed to be the effect of a fading of 'intellectual scars' induced by bad home conditions, and occurred more as a result of *removal from* such conditions rather than of entry into relatively better ones. Those with the greater damage due to deprivation made the greater recovery (often many years later) and they made it more rapidly.

Social changes

A significant social factor is reflected by the difference between the two groups in terms of present availability; at the end of 1955 only nine of the original twenty-five from very bad homes were still available, whereas a larger proportion, nineteen out of thirty-four, from less adverse conditions were still under the care of the hospital. Since many of the remainder had been discharged, this adds support to the generalization that, in this type of population, the worse the social background, the better the prognosis from all points of view. Moreover, of the nine remaining from the original group from very bad homes, eight were already working in the community although still under certificate, and one was receiving industrial training prior to community placement. Only four of the nineteen remaining from the original group from less adverse circumstances were in a similar position at the end of 1955.

It is unlikely that I.Q. increments play a major part in *causing* improved social adaptation, since the correlation between I.Q. and social competence at this level is not very high. These increases must therefore be regarded as one facet of a process of more or less total personality development which is also likely to be associated with the same general causation.

III. Second Study

Cognitive changes

The second study followed up all patients of feebleminded grade first admitted to this hospital in 1951 who were still available at the beginning of 1956. The purpose of this investigation was to find out whether, among other things, I.Q. changes in this population and age range were a direct function of time interval between test and retest, as is sometimes implied for normals (cf. Eysenck, 1953). Retesting these subjects in 1955–6 gave

an interval between tests of 4½ years, roughly double that of the *original research*; it was thus possible to ascertain whether the average increments would also be doubled. Of an original sixty patients admitted in 1951, thirty-two were still available, and, as before, it seemed likely that they were the less promising members of the original sample.

The following methodology was employed. An independent investigator from another hospital examined in detail the thirty-two case-histories, using the twelve criteria evolved in the *original research* (see footnote, p. 145) and divided them into a group from very adverse homes and a residual group from less adverse conditions. As before, this latter contained many from bad homes, though not exceptionally bad, and very few from good homes. The external investigator knew neither the patients nor their test scores. It was then predicted that members of the former group would tend to make the larger increments. Subsequently, the retesting was carried out by one of us (S. R.) who was ignorant of the patients, their social histories, the group to which they had been allotted, the prediction which had been made, and their earlier test scores. The results are shown in Table 2.

Table 2. *Patients admitted in* 1951, *retested* 1955–6

	Group from very bad homes			Group from less adverse homes		
	Wechsler Full Scale I.Q., 1951	Wechsler Full Scale I.Q., '55/'56	Diff.	Wechsler Full Scale I.Q., 1951	Wechsler Full Scale I.Q., '55/'56	Diff.
Mean	63·1	74·6	11·5	69·0	76·2	7·2
S.D.	15·0	15·2	6·3	13·9	14·6	6·6
N	13	13	13	19	19	19
t (1-tail test)	—	—	6·894	—	—	4·926
Significant at	—	—	0·1%	—	—	0·1%
I.Q. ranges:	1951, 37–90; 1955–6, 54–99			1951, 42–94; 1955–6, 55–101		
Age at retest:	Mean 23·0, S.D. 5·9 years			Mean 23·6, S.D. 6·1 years		
Period between test and retest:	Mean 53·6 months, S.D. 3·8 months			Mean 51·6 months, S.D. 5·0 months		
Range of I.Q. changes:	0 to +18 points. (10 points and above: 9/13 = 69%; 15 points and above: 7/13 = 54%.)			−7 to +20 points. (10 points and above: 8/19 = 42%; 15 points and above: 1/19 = 5%.)		

t (between group from very bad homes and group from less adverse homes) = 1·925, significant at the 5% level (1-tail test). This difference is much less significant than in the *original research*, and probably arises from a combination of four factors: (1) Total *N* is much smaller. (2) One patient in the group from less adverse homes showed a very atypical gain of 20 points, which alone was responsible for depressing statistical significance from the 2½% to the 5% level; at the time of first testing it had been noted that she was unusually disturbed emotionally, and came from a background, which though not quite conforming to the necessary criteria, was far from adequate. (3) The only patient in the group from very bad homes who did not change at all was later found to have been suffering from a right temporal malignant tumour during the material period. (4) Selective wastage had almost certainly occurred since 1951, since rehabilitation and discharge rates have accelerated greatly since then. Consequently, there has probably been a tendency for the better members of the group from very bad homes to have left the institution.

It will be seen that once again the subjects from very bad homes tended on the average to make the larger increments, although in this sample the discrepancy between the *average* increments by the two groups is less marked than in the *original research*. With roughly double the original time interval, it is clear that, on the one hand, the present group from very bad homes did not increase greatly their average increment compared with that achieved by the original group (mean increment 11·5 compared with 9·7 points). Clearly, therefore, the amount of gain in this group is not a direct function of the length

of time between test and retest. On the other hand, it will be seen that members of the residual group from less adverse conditions almost doubled their mean increment over the doubled time interval (mean increment of 7·2 compared with 4·1 in the *original research*). Thus their gains are much more a direct function of the test-retest time interval. This is consistent with findings in the present *First Study* in which the maximal gain for the group from very bad homes occurred in the first 2–3 years, with a relatively slow increase thereafter, while those from less adverse conditions made a steady increase over the 6-year period. Thus, the difference between the gains made by both the groups is maximal at the beginning of the period in which increments have been studied, and subsequently because the growth rate of the bad home group tends to decelerate sharply (see Table 1), while the residual group maintains a steady progress, the difference between the average gains decreases.

Social changes

The present sample on the average had been under care of this hospital a shorter time than any of the others and was younger. Nevertheless, social changes had already begun to be apparent, and differences, as yet small, between the two groups were noticeable. Thus seven out of the thirteen from very bad homes were either undergoing intensive training prior to community placement (see next Study) or were already in outside employment, while only six out of the nineteen from the less adverse homes were similarly placed.

IV. Third Study

This study was devoted to a special problem; the *original research* had suggested that I.Q. changes resulted as an effect of removal from very adverse conditions rather than from entry into relatively better ones; that is, absence of such unfavourable circumstances enabled intellectual scars to fade. An important problem then envisaged was to determine whether special environmental conditions would aid this apparently natural process of recovery. A group of patients was readily available for the testing of this hypothesis.

For some years the Psychology Department has been responsible for running two small rehabilitation units for selected high-grade patients. These are taught industrial skills in small workshops. develop good work habits, receive educational aid and experience of the community, eventually being placed in local industry at standard wages and in normal jobs. During their initial period outside they return to the hospital nightly but after a time may graduate to lodgings and obtain their discharge. Observation, and their success rate, suggested that many undergo radical changes under these conditions, and the question thus became whether these involved intellectual aspects of personality, and, if so, whether these gains were environmentally induced. The scheme briefly mentioned has been described in detail elsewhere (1954*a*).

At any one time about twenty to twenty-five patients so trained are at work in the community, and these formed the subjects of the present experiment. All had at some time during the last few years been given Wechsler, Form I, and this enabled retest results to be compared with their earlier scores. During the period between test and retest all had had at least 6 months intensive training, and at least 6 months in the community (and often more), both of which involved considerable change of circumstances for these persons. Social histories before admission were again assessed by the external investigator

who knew neither the patients nor their test scores, and retesting was carried out by one of us (S.R.) who knew neither the patients, nor their early histories, nor their earlier test scores. Table 3 shows the relevant data for both groups.

The first point of note is the rather higher proportion than usual of the whole group (a little over half, compared with about 42% in a more random sample) who came from very bad homes; this is likely to be the result of a selection factor, namely that such persons tend to do better socially than others and are therefore more likely to be chosen for special training and rehabilitation. Once again, the larger increments tended to occur in these patients, although the difference did not reach an acceptable level of significance; this may have resulted from a selection factor favouring the more promising patients

Table 3. *Wechsler full-scale test-retest results on persons specially trained in rehabilitation units and placed in the community in normal employment*

	Group from very bad homes			Group from less adverse homes		
	Test I.Q.	Retest I.Q.	Diff.	Test I.Q.	Retest I.Q.	Diff.
Mean	64·8	78·9	14·1	75·1	85·5	10·4
S.D.	13·7	13·1	6·0	13·4	12·4	6·3
N	11	11	11	10	10	10
t (1-tail test)	—	—	7·772	—	—	5·547
Significant at	—	—	0·1%	—	—	0·1%
I.Q. ranges:	First test, 49–94; Retest, 64–105			First test, 63–108; Retest, 68–112		
Age at retest:	Mean 24·2, S.D. 3·9 years			Mean 27·6, S.D. 6·9 years		
Period between test and retest:	Mean 54·6 months, S.D. 17·6 months			Mean 47·0 months, S.D. 17·6 months		
Range of I.Q. changes:	+5 to +20 I.Q. points. (10 points and above: 9/11 = 82%; 15 points and above: 7/11 = 64%.)			+2 to +19 I.Q. points (10 points and above: 5/10 = 50%; 15 points and above: 3/10 = 30%.)		
Time in training:	Mean 10·6 months, S.D. 6·2 months			Mean 12·4 months, S.D. 5·2 months		
Time in employment:	Mean 13·2 months, S.D. 10·4 months			Mean 16·4 months, S.D. 11·7 months		

t (between group from very bad homes and group from less adverse homes) = 1·449, not significant (1-tail test).

from the less adverse conditions. While in both groups the average increase is rather large, it is not so very different from what would be expected under ordinary hospital conditions. We are able to compare these data with the results of the *Second Study*, Table 2. The average test-retest time interval is similar, and in the *Second Study* the average gain of those from very bad homes was 11·5 points, S.D. 6·3, while in the present study the similar group increased on the average by 14·1 points, S.D. 6·0. Similarly, the group from less adverse homes in the *Second Study* made an increment of 7·2 points, S.D. 6·6, compared with an average of 10·4, S.D. 6·3, from the similar group in the present investigation. The average difference between the gains in the *Second* and *Third Studies* is thus a rough measure of the effect of special environmental conditions; for the groups from bad homes the difference is 14·1 minus 11·5 = 2·6 points, and for the remaining group 10·4 minus 7·2 = 3·2 points. These are then the maximum effects which can be ascribed to environmental difference, and they are clearly small.

As has been stated, subsequent I.Q. increment seems to occur as a long-term effect of removal from a traumatic environment, and there can be no doubt that the earlier intellectual deficit resulted, partly at least, from *negative* environmental effects in childhood. The present study has indicated that fairly clear environmental differences in this age-group do not exert much of a *positive* effect upon the I.Q. There are two possible explanatory factors. First, the difference between the ordinary institutional environment

and the one provided by intensive training in special units followed by community experience may not be as large as thought. On the other hand, it must be noted that the difference is certainly large enough to produce considerable social changes, resulting in greater success and more rapid discharge, so that there is little doubt that the environment is a specially stimulating one.

Secondly, it is possible that the findings reflect both an initial *negative* effect in childhood and a *masked positive* effect in adolescence and early adult life. It may be that any environment, other than an adverse one, will stimulate the apparently natural recovery process, and that the effects of environmental differences are small simply because all—even the institutional environment—have a *positive* effect. That environmental effects cannot themselves be measured needs stressing; only the effects of environmental differences are measurable. Precisely this same point has been made about the classic studies of identical twins brought up apart; in only a few cases was each twin in a radically different environment and thus the variable 'environment' was seldom accessible for full measurement.

Another source of evidence bearing out the present conclusions emerges from a comparison of the results of the *First Study* with this one. The time intervals are not quite similar, but it will be noted that the members of both groups from very bad homes made very similar mean increments. Eight of the nine members of this group in the *First Study* were employed in the community in jobs traditionally provided for the mentally deficient and for which they had no special training; yet their I.Q. gains were closely similar to those in the present study. So, too, with the members of both groups from less adverse homes; those who had had no special training and the majority of whom had had no experience of outside employment (*First Study*) made similar increments to those who had had both. There was no doubt, however, that special training and normal experiences had induced profound social changes in the patients of the present study. They were all holding good jobs with wages ranging from £3. 0*s*. 0*d*. to £8. 2*s*. 0*d*. per week with an average of £5. 17*s*. 0*d*. All were expected to gain their discharge, and to become reasonable and self-supporting citizens.

Summing up, it seems clear that intellectual retardation or arrest can and does occur in conditions really antagonistic to mental health in childhood. And it is precisely these conditions from which the vast majority of the certified feebleminded are drawn. Entry into any better environment seems equally able to stimulate increments in I.Q. and to allow the damage to fade. This, however, is a long-term process which will be discussed in detail in the next section.

V. Discussion

(*a*) *Variability of the* I.Q.

At this stage, brief discussion of the causes of I.Q. variability is appropriate. Spurious changes can be subsumed under the first three of the following headings. (1) *Errors of measurement* which in most cases represent *personal fluctuations* (e.g. alterations in mood, misunderstanding of the instructions, chance decisions, fatigue, boredom, anxiety and changes in motivation on the part of the subject, and errors in scoring, instructions, or in the establishment of rapport on the part of the tester). These combine to reduce test-retest correlation from a theoretical +1·0 to an average of 0·9 on immediate retest

Thorndike, 1933). Similarly, standardization differences (e.g. varying S.D.'s on the 1937 Stanford–Binet at different ages) will, if uncorrected, produce spurious changes. (2) *Test practice and coaching* have well-known effects (see Vernon, 1954). (3) *Incorrect testing* by persons unqualified in psychometrics may clearly produce errors resulting in excessive I.Q. variability upon retest. (4) *Variation in the rate of intellectual growth.* In general, as the interval between tests lengthens, so the retest correlation decreases, often steadily (Nemzek, 1933; Thorndike, 1940; Eysenck, 1953). There is no reason to suppose that errors of measurement (whether arising from personal fluctuations in respect of subject or tester, or standardization errors) are likely to be steadily cumulative, and thus a decreasing correlation coefficient must be primarily a reflexion of some real change of status of members of the group.

In this connexion, it is necessary to discuss 'test reliability' since confusion tends to arise over this concept. Test reliability relates to two factors: first, whether the test measures the same quality in the same way on different occasions; and secondly, whether the quality measured itself alters. The first point has been discussed under 'errors of measurement', and the second might well be termed 'unreliability of persons' since, as we have seen, the fact of decreasing correlation with increase of time implies just that. It is this in particular which produces difficulties; it is by no means uncommon to hear or read the following tautological argument: 'over several years test results change because of imperfect correlation or test unreliability'. It is further implied that such changes are statistical, reside in the test rather than in the person tested, and therefore have no psychological significance. Obviously, however, the test unreliability or lowered correlation over several years is more likely to be a reflexion of change of status of individuals rather than vice versa.

Statistical regression towards the mean is another concept frequently used to account for I.Q. changes; where this concerns errors of measurement on the same or different tests, or where different tests have a low initial correlation when given within a day or two of each other, then this is entirely correct. In such cases, regression of extreme scores is inevitable and of no psychological significance. In a cogent discussion of regression, McNemar (1940), agreeing that it is a descriptive rather than an explanatory concept, points out that an inferior or superior group will not move towards the general population mean on retest unless it has been specifically selected on the basis of initial test results (e.g. Terman's group of gifted children, selected on the basis of I.Q. 140 and above). Within a group not so selected, regression effects will occur around the group mean, be it superior or inferior. We have already seen that, granted proper standardization and test reliability (i.e. small errors of measurement) a steadily decreasing correlation coefficient tends to occur with increase of time, such that after 5 years it may be about 0·7. This implies some changes of as much as 15–20 points, and it would be quite incorrect to say that these are 'due to regression' with consequently no psychological meaning. Yet such statements are not uncommonly made, and in effect there appears to be a tendency to regard statistical expressions as somehow primary to behavioural data, with curious results. Regression cannot explain such large changes which reflect primarily an alteration in response to the test situation; naturally the implications of such variations are not necessarily obvious and must be interpreted with caution.

At the risk of labouring this point, let us see why in the *original research* and the present studies, the concept of regression cannot explain the major part of the increments.

First, the groups were not chosen for study or follow up on the basis of initial test score. Secondly, the *original control group* retested after a short time interval made an average increment of only 4 points, which must be a maximal estimate of errors of measurement including regression effects for this type of sample. Stating the same point differently, the Wechsler test has a very high immediate retest reliability, hence regression effects must be small. Thirdly, such wastage of subjects as occurred in the *First* and *Second Studies* was for reasons other than initial test results. Fourthly, the two groups which formed the main sample in the *original research* were initially equal in I.Q. Yet by the time that investigation was complete, and even more so now, different cognitive and social trends had occurred, and regression cannot account for such differences in initially equated groups. Fifthly, statistical regression is psychologically meaningless. Throughout all our studies, however, there has been a consistent tendency for the amount of 'regression' to be related to psychologically meaningful external criteria, the record of early experiences. It is hardly likely that the purely statistical would be so linked with the purely behavioural. Thus on all counts, statistical regression cannot explain the major portion of the I.Q. increments.

(*b*) *The present findings*

The increments noted both in the *original research* and in the present studies are likely to have been largely the effect of 'unreliability of persons', that is, of variations in the rate of late intellectual maturation. All four investigations underline the better prognosis for the feebleminded person from bad social conditions; they tend to have larger I.Q. increments, to enjoy a greater success within the hospital, to be selected for intensive training by virtue of their other qualities, to respond to training a little more rapidly and to be tried in the community earlier and more often than those from rather less adverse social conditions. They also tend to achieve discharge from care more rapidly and more often, and although sometimes they may have a lower I.Q. in adolescence than the other patients, they tend to catch up and overtake them within a few years. Indeed they present a substantially different picture in many spheres after, say, 6 years, and often these changes occur at relatively late ages. In general, these results are strongly supported by the social findings of Charles (1953) in which a very long-term follow-up, conducted with exemplary care, of children originally judged mentally deficient, showed that they had not developed according to a rigid stereotype. Many at the ages of 30 and above were in the normal or dull normal range and were holding their own in the community. The outcome for the majority was far better than would have been predicted at the time of their attendance in the 'opportunity rooms' of Lincoln, Nebraska, schools. Similarly, any person reading the case-histories of many of the present subjects cannot fail to be struck by the discrepancy between the gloomy predictions of psychologists and psychiatrists made within the last 10 years, and the actual progress made. Delayed personality maturation may well be a factor explanatory of Penrose's (1949) observation that the incidence of high-grade deficiency decreases sharply from the age of 15 onwards.

What, then, are the precise mechanisms of these changes? Earlier papers (1954*b*,1955) argued that at all stages the obtained I.Q.'s were relatively accurate measures of the then intellectual status of the subjects, rejecting the hypothesis of pseudo-feeblemindedness, which, as will be seen, really begs the question. The work of Hebb (1949) is clearly relevant; adverse conditions in early life would be likely to interrupt or seriously damage the for-

nation of 'phase sequences', and the more adverse these conditions the greater the eficit to be made good. It appears that improvement may not commence until adoescence and may continue well into the twenties. In many cases there thus appears to be period of latency, during which growth rates may show little acceleration, lasting from hildhood and infancy when the damage occurred, until adolescence. We possess far less ata on this period, however, and this view is inferred from several sources: (*a*) the few hildren studied with similar backgrounds to those of our older subjects appear to change ttle; (*b*) the case-histories of our subjects do not imply, in most cases, a level in childood very much inferior to their status in adolescence; and (*c*) most of the early investiators believed that intellectual growth in the feebleminded ceased several years earlier han in the normal. Our own results suggest precisely the opposite for many such ersons, but it is a reasonable assumption that this original view about intellectual levelopment in the feebleminded sprang from the fact that children were mainly being tudied at that time, and it may be that over late childhood there is a developmental lateau which, when it begins, suggests a cessation of growth. While this latency period s an hypothesis, it seems to be a reasonable one, and at all events it seems quite certain hat in childhood these persons do not change as rapidly as later. Removal of most eebleminded children from traumatic conditions usually took place years before entry nto a mental deficiency hospital, the interim period being spent in residential E.S.N. chools, children's homes or orphanages. Then, perhaps at age 16, certification took place s the last link in a long chain of official intervention.

All one can say in explanation of this hypothetical latency period is that it may be eeded to make good the foundations upon which subsequent development depends, nd to establish or re-establish some integration of personality—that perhaps 5 or 10 years re needed to give the developmental basis normally gained during the first few years of ife. Certainly future investigations must be concerned with studying such persons at a much earlier age, and as soon as possible after removal from adverse homes.

A further and necessary hypothesis requiring some discussion is that there must be imits to the change which occurs; these limits may be set by the basic resources of the entral nervous system, genetically determined, and also, perhaps, by the amount and quality of the damage produced by negative environmental influence. The largest increment ever noted in this hospital on the same test was 29 points but this was over about years; this then appears to be the maximum over a similar time period in adolescence nd early adult life among such subjects. Over the whole of a person's life span, however, he limits to change may be slightly greater, say, a maximum of 40 points (cf. Vernon, 955); obviously, however, the majority of the feebleminded may never make changes f this order. Not only do the amounts of change differ in different subjects but also he rates of change. We have seen in the *First Study* that those from really adverse onditions tended to make a large initial gain over about 3 years, which on the average vas halved during the next 3 years; those from less adverse homes (but nevertheless bad nes) tended to make a steady gain over the 6-year period. The first interpretation nentioned was that the rate of recovery depended upon the severity of initial damage. The concept of limits may, however, also provide an explanation. It may be that the nembers of the group from very bad homes were reaching their limits of change and herefore growth rates were decelerating. Longitudinal research over about 10 years night elucidate this important point.

The main implications of the cognitive and social findings reported here seem clear; we cannot predict a necessarily poor outcome for children with I.Q.'s in the 50's, 60's and 70's if they also come from very adverse environments. While there is no doubt much truth in Bowlby's thesis that separation from the mother may have very unfortunate effects on the child, we must not interpret this finding too rigidly for, as Hilda Lewis (1954) points out, it may be even more disastrous to leave some children with unsuitable parents. Moreover, findings such as these suggest that feebleminded persons, most of whom come from bad or very bad conditions, are within limits far from being the hopeless propositions which until recent years was generally accepted; already a much more positive attitude towards them is emerging.

(c) *Relevance of other work to the present findings*

That environment can exercise a stimulating or depressing effect upon intellectual function is now generally accepted (cf. Vernon, 1955), although there is still a dearth of knowledge about precise mechanisms, the limits to, and the factors limiting, such changes. So far as the mentally deficient are concerned, however, the concept of pseudo-feeblemindedness continues often to be used to explain away I.Q. increments among them. Perhaps the best statement of this view is that of Porteus (1941): 'Very wide differences in intellectual status merely indicate that the first diagnosis was wrong. Any child who finally functions at a normal level proves thereby that he never was feebleminded.' In effect, Doll's attitude (1941) is similar; essential incurability is suggested as a necessary criterion of mental defect; one might, therefore, have to wait 20 years to find out whether a child improved and hence no firm diagnosis of a high-grade case could be made in childhood. Both writers express an outlook firmly rooted upon the concept of a fairly rigid constancy of the I.Q., a view to which few would now subscribe. Elsewhere, (1955) we have argued that the concept of pseudo-feeblemindedness is now untenable in cases showing deferred maturation, since most studies show that initial testing gave no indication of underestimation at the time, and must be considered a true picture of the then present status of the person. Thus Porteus's view that *mistaken diagnosis* has occurred must be rejected; rather, the clinician has made a *mistaken prognosis*. And the present studies have shown that it is possible to predict I.Q. changes by an external criterion. It is important to realize, therefore, that arrest in intellectual development, or a decelerated growth rate, may be permanent or impermanent to varying degrees. In both types, the symptoms (e.g. social incompetence, behaviour disorder, or low I.Q.) may be similar, but it is a truism that the same symptom may arise from different causes.

Previous researches on I.Q. changes in the mentally deficient have been critically reviewed elsewhere (Clarke & Clarke, 1958). Here, it is proposed to discuss only one specific study relating to English feebleminded patients, and two others with broader implications.

(1) Mundy (1955), in studying environmental influences on the I.Q., used an experimental and control group each consisting of twenty-eight adult feebleminded women, and believed that the two groups were initially equated on a number of important variables. The interval between test and retest was 2 years, during which time the controls remained within a good institution (a convent) while the members of the experimental group enjoyed normal life experiences, mostly in residential wage-earning employment. On

retest, the control group showed an average increment on the Wechsler test of 2·25 points, and on the Progressive Matrices of 3·46 points. The average gains made by the experimental group, however, were 11 and 9·82 I.Q. points, respectively. Mundy believed that these results indicated that present environmental differences induced differential gains, but the validity of this conclusion clearly rests upon whether control and experimental groups were in fact *initially equal in all respects*. We are told, however, that there was no clinical selection by the hospital for outside employment, and that 'any able-bodied patient who wished for outside work was found suitable employment as soon as this was feasible'. This suggests that the members of the experimental group had all, through drive, ambition and initiative, asked for outside employment and hence gained it, while for several years at least the members of the control group had been content to remain within a good institution: there was thus a difference between the groups in terms of 'self-selection'. It should be added, however, that in a personal communication the author has stated that some cases were in effect originally transferred to employment for administrative reasons; this would reduce the amount of 'self-selection' but it might well be an indication of selection by the hospital. A further reason for believing that the control group contained atypical patients was the extraordinary stability of their test scores over the 2-year period; twenty-five out of the twenty-eight remained within 0 and plus 4 I.Q. points of their initial Wechsler result, and no patient increased by more than 6 points. This would imply a much higher test-retest correlation than is reported elsewhere in the literature. Because of these two factors, therefore, Mundy's conclusion is open to some doubt, and, in view of our own findings, it may well be that the I.Q. changes would have occurred differentially in any reasonable environment. Where her research and our own agree, however, is that early very adverse experiences played a part in retarding intellectual and general development. (See also Mundy, 1957.)

(2) Some striking parallels with our own findings emerge from experimental work in an entirely different field of deprivation, although these analogies ought not to be pressed too far. Widdowson & McCance (1954) studied the effects of different food diets on undernourished German orphanage children. Among other things, their growth in height, weight and ossification were investigated. The most striking observation was that following the period of undernourishment, these children grew even better than normal children of the same age, in spite of the very simple diet provided, which contained little milk or animal protein, and up to 75% of its calorie value in the form of flour products. It may be either that the flours of different extraction rates were equally nutritious or that the differences between them were masked by the exceptionally rapid growth rates of undernourished children when receiving a diet adequate in calories. Even the addition of a special milk supplement did not increase further these already accelerated rates of increase. In an appendix to this report (*op. cit.* p. 80) J. O. Irwin indicated that alternative viewpoints are equally possible. Because of original underfeeding, these children might have grown on any diet not grossly inadequate, the response to different amounts of vitamins would therefore be largely non-specific, and specific differences might still be detected in children with a more normal nutritional history on similar diets. On the other hand, others might be more struck with the fact that these children grew equally well on all the diets and at a faster rate than normally. The similarity between these concepts and those relating to our own study of environmental effects is very close. Another interesting finding, so far as weight increases were concerned, was that initial increases

were more rapid than later ones (cf. our *Second Study*). It is of some consequence that unusual growth rates (with respect to normals) after malnutrition should in some ways be analogous with unusual increments in cognitive and social development in persons who suffered a different form of deprivation.

(3) The work reviewed by Bowlby (1951) is too well known to need summary; it is highly relevant to our thesis that early adverse environmental influences have played a part in producing mental deficiency. His monograph is, however, almost entirely concerned with studies of children or young adolescents. In general, Bowlby is pessimistic about their recovery from early psychological damage, particularly in the absence of psychotherapy. Yet our own subjects, the majority of whom could scarcely have suffered worse childhood experiences (followed by institutional upbringing in many cases) being therefore *par excellence* a deprived group, have on the whole done much better than might have been expected, and none has received formal psychotherapy. Similarly, some of the social studies such as that of Charles (1953) have indicated that persons who were originally high-grade defectives in childhood do not develop according to a rigid stereotype, that a wide range of individual differences is maintained, with many eventually functioning in adult life at dull-normal or normal levels. There is no way of knowing what our subjects would have been like had they enjoyed normal conditions from birth onwards, but there is no doubt that in adult life they are, on the whole, very much less impaired than would have been expected from Bowlby's original theory.

Two very important general questions arise. First, to what extent are the main variables those of human susceptibility to early psychological damage and resilience thereafter, and what is their nature? Why is it that some children are utterly overwhelmed by adverse conditions while others are apparently stimulated by them? We have no real estimate of the proportion which emerges relatively unscathed, simply because we tend only to study the problem cases, but it may well be larger than we think. For example, often two or three children in a family come to the notice of the authorities while their other siblings do not, and we know next to nothing about the latter. Secondly, is it possible that, when more deprived children are followed up through adolescence and adult life, it will be found that, in many, the initial psychological damage, as with our own deprived population, tends steadily to be repaired to varying extents? This, if true, would represent a need for modification of the 'maternal deprivation' thesis, and it is of great interest that Bowlby, Ainsworth, Boston & Rosenbluth (1956) have already begun to re-formulate some aspects of it.

The writers are greatly indebted to Dr J. F. MacMahon, Physician Superintendent, for his many suggestions and constructive criticism; to Dr Gordon Claridge for having made independent assessments of social histories; and to Dr H. R. Beech for statistical advice and for the computations.

References

Bowlby, J. (1951). *Maternal Care and Mental Health.* Geneva: World Hlth Org.

Bowlby, J., Ainsworth, M., Boston, M. & Rosenbluth, D. (1956). The effects of mother-child separation: a follow-up study. *Brit. J. Med. Psychol.* **29**, 211–47.

Charles, D. C. (1953). Ability and accomplishment of persons earlier judged mentally deficient. *Genet. Psychol. Monogr.* **47**, 3–71.

Clarke, A. D. B. & Clarke, A. M. (1953). How constant is the I.Q.? *Lancet*, ii, 877–80.

Clarke, A. D. B. & Clarke, A. M. (1954*a*). A rehabilitation programme for the certified mental defective. *Ment. Hlth, Lond.*, **14**, 4–10.

Clarke, A. D. B. & Clarke, A. M. (1954*b*). Cognitive changes in the feebleminded. *Brit. J. Psychol.* **45**, 173–9.

Clarke, A. D. B. & Clarke, A. M. (1955). Pseudo-feeblemindedness—some implications. *Amer. J. Ment. Defic.* **59**, 507–9.

Clarke, A. M. & Clarke, A. D. B. (Ed.) (1958). *Mental Deficiency: the Changing Outlook.* London: Methuen. (In the Press.)

Doll, E. A. (1941). The essentials of an inclusive concept of mental deficiency. *Amer. J. Ment. Defic.* **46**, 214–19.

Eysenck, H. J. (1953). *Uses and Abuses of Psychology.* London: Penguin.

Hebb, D. O. (1949). *The Organisation of Behavior.* London: Chapman and Hall.

Lewis, H. (1954). *Deprived Children.* London: Oxford University Press.

McNemar, Q. (1940). A critical examination of the University of Iowa studies of environmental influences upon the I.Q. *Psychol. Bull.* **37**, 63–92.

Mundy, L. (1955). Environmental influences in intellectual function as measured by intelligence tests. London University, unpublished M.Sc. Thesis.

Mundy, L. (1957). Environmental influence on intellectual function as measured by intelligence tests. *Brit. J. med. Psychol.* **30**, 194–201.

Nemzek, C. L. (1933). The constancy of the I.Q. *Psychol. Bull.* **30**, 143–68.

Penrose, L. S. (1949). *The Biology of Mental Defect.* London: Sidgwick and Jackson.

Porteus, S. D. (1941). *The Practice of Clinical Psychology.* New York: American Book Co.

Thorndike, R. L. (1933). The effect of the interval between test and retest on the constancy of the I.Q. *J. Educ. Psychol.* **24**, 543–9.

Thorndike, R. L. (1940). 'Constancy' of the I.Q. *Psychol. Bull.* **37**, 167–86.

Vernon, P. E. (1954). Symposium on the effects of coaching and practice in intelligence tests. V. Conclusions. *Brit. J. Educ. Psychol.* **24**, 57–63.

Vernon, P. E. (1955). The assessment of children. In *Studies in Education*, no. 7. University of London Institute of Education.

Widdowson, E. M. & McCance, R. A. (1954). Studies on the nutritive value of bread and on the effect of variations in the extraction rate of flour on undernourished children. *Spec. Rep. Ser. Med. Res. Coun., Lond.*, no. 287. London: H.M.S.O.

(*Manuscript received* 14 *March* 1957)

SHAPE RECOGNITION: A REPLY TO DEUTSCH

By P. C. DODWELL

Psychology Department, Birkbeck College, London

The main points made by Deutsch (1958) call for some comment. First, there is ample evidence in Lashley's (1938) reports to show that his rats were *in fact* discriminating by a part of the visual display in most cases and not by the whole; this would seem to indicate that the rats were not basing discriminations on the properties of complete shapes, whether those shapes could be 'seen whole' or not. Admittedly, putting the argument in terms of the rat's (inferred) field of distinct vision is weak, but this does not affect the fact that Lashley demonstrated a small field of *effective* vision in his experiments. Secondly, there is some evidence to show that rats do not treat rotated shapes as equivalent, although they may learn to do so. Fields (1932), in his investigation of the rat's ability to discriminate a triangle in 24 different positions (rotations) from a circle, found that the average number of trials needed for the initial discrimination was 580; for the first rotated position, 230; for the next 11 positions, an average of 45 trials each; and for the remaining 11 positions, 21 trials each. This suggests that the rat may be learning to treat a family of related signals in the final common cable (such as would be generated in the writer's system) as equivalent to one another. It is difficult to see why such learning should be necessary on Deutsch's theory, since the messages for all positions of rotation would be the same. Furthermore, on pairing a square or rectangle (base horizontal) with a rotated triangle, the animals tended to choose the shape with horizontal base in preference to the triangle. In the writer's theory such a 'wrong' choice is explicable, since the signal for the square or rectangle has one feature in common with the signal for the first position of the triangle—a long delay on the horizontal f.c.c. attributable to a horizontal base. In Deutsch's theory such a choice could not be explained. This evidence is not conclusive, however, since Fields did not demonstrate whether his rats were responding to parts of the display or the whole as Lashley did. In Fields's experiment, the visual angle subtended was approximately 18°.

Deutsch says that there is evidence that rats jump spontaneously to horizontal bars, but does not quote an experiment which shows this. It was not the case in the writer's experiment, although the horizontal/vertical discrimination was quickly learned. Deutsch suggested a plausible reason why Lashley's rats could discriminate a square rotated through 45° from a circle but not a square from a circle, and this same argument *could* explain why they discriminated a square from another rotated through 45°. Indeed, if Deutsch's theory holds, such an explanation (or something similar) would be needed. Gellermann's (1933) finding is open to the same interpretation as that given above for Fields: the signal for the rotated triangle is 'recognized' as belonging to the same family as the signal for a triangle with base horizontal, and the receptors are rotated to confirm this. In Deutsch's theory it is difficult to see what purpose such a rotation would serve.

It is true that Deutsch can incorporate a system for separate coding of information about tilt in the manner he suggests; however, the evidence seems to point to a coding

system (at least in visually primitive organisms) in which information about shape and tilt are compounded in one signal, or give rise to signals which mutually interfere with each other, which is not the case in Deutsch's (modified) theory.

References

Deutsch, J. A. (1958). Shape recognition: a reply to Dodwell. *Brit. J. Psychol.* **49**, 70–1.

Fields, P. E. (1932). Studies in concept formation. I. The development of the concept of triangularity by the white rat. *Comp. Psychol. Monogr.* **9**, 1–70.

Gellermann, L. W. (1933). Form discrimination in chimpanzees and two-year-old children. *J. Genet. Psychol.* **42**, 3–27.

Lashley, K. S. (1938). The mechanism of vision. XV. Preliminary studies of the rat's capacity for detail vision. *J. Gen. Psychol.* **43**, 123–93.

(*Manuscript received* 4 *May* 1957)

CRITICAL NOTICE

By JEROME S. BRUNER
Harvard University

Thinking: An Experimental and Social Study. By Sir Frederic Bartlett. London: George Allen and Unwin. 1958. Pp. 203. 18*s*.

Sir Frederic Bartlett is a man remarkable for the combination of common sense and high adventure that has always informed his work. The interplay of shrewd conservatism and speculative daring in this, his most important work since *Remembering*, reminds one of nothing so much as the blend of loyalism and freebooting in those earlier Elizabethans, the privateers. He captures prizes, but he loses not his head or his calm. This is a remarkable book, 'vintage Bartlett' as Sir Frederic's students would say, full of insights and ambiguities. If it does not succeed in taming its subject, *thinking*, it certainly leaves it well worked over. A reviewer may perhaps be forgiven for introducing a book in so biographical a view; for this is in such a deep sense a biographical book. Not only is the reader witness to a dazzling display of uniquely Bartlettian intellectual technique, but the author offers among others his own past efforts (the construction of *Remembering*), as laboratory material on which to try out his notions about the process of 'adventurous thinking'.

The point of view, the theoretical line of the book finds its origin in the analogy between skilled bodily performance and the process of thinking. Sir Frederic does not insist on a parallel between the two, but rather sees the latter as an extension of the former, one made possible by the capacity of man to manipulate symbols rather than being bound to the employment of receptor-effector acts. Both bodily skill and effective thinking depend upon a fitting use of evidence, the former requiring sensory regulation, the latter something more difficult to define. Each is marked by a sequential integration of component processes, and in consequence, each can be shown to exhibit certain common or analogous properties. One of these is *timing* in the use of information and in the executing of steps, be they muscular acts or acts of inference. Each type of sequence, moreover, has its *stationary phases*, its critical 'hoverings' and 'halts' during which timing is regulated. Betimes, each sequence reaches a 'point of no return', the point at which the batsman can no longer change his stroke without the price of disruption, the point at which the thinker becomes the victim of what he has done to the evidence with which he has had commerce. Throughout, each process is pregnant with its terminus; moving toward an issue—be it the drawing of a conclusion or the delivery of a drop shot. This is the '*direction*' of the sequence, and upon it Prof. Bartlett rests much of the theoretical edifice he constructs.

Since, I believe, the book gets into some difficulties by relying on the parallel between bodily skill and thinking, a word needs saying about the difference between the two, and how it is that Prof. Bartlett conceives the one developing from the other. One finds the clue to the difference in a close examination of the nature of 'direction'. In its simplest form, direction is identified with 'receptor-effector anticipation' in which perceptual scanning runs a bit ahead of the action to be performed with the result that reaction has a seemingly preternatural swiftness. If, for example, we are to move toggle switches in the direction indicated by series of independent and discrete displays, the time required for the making of a movement is absurdly little less than the time it takes the operator merely to observe or interpret what the display is saying. But this is a narrow range of anticipation, very small beer. Bartlett offers a second stage en route to thinking, 'perceptual anticipation', in which the operator deals not with single and independent events but with the directional properties of sequences—with sequential redundancy, if you will. Such sequential anticipation is possible *provided* the operator has created a model or schema to represent the recurrent features of the sequence, a problem of learning and information storage toward which Prof. Bartlett is oddly cavalier. The author suggests that 'perceptual anticipation comes into the picture because the behaviourally more basic receptor-effector anticipation very often breaks down'. If this is true, then it should also be true that schematizing in memory occurs because efforts to deal with single items by rote 'very often break down'—a view to which the author of *Remembering* should take violent exception, even if it is stated by the author of *Thinking*!

Finally, 'when we discover or learn how to use warning signals or symbols' in dealing with the evidence

provided by sequences of events, then we are dealing with the kind of direction that is most characteristic of thinking.

What, then, is thinking? 'The process begins when evidence or information is available which is treated as possessing gaps, or as being incomplete. The gaps are then filled up, or that part of the information which is incomplete is completed. This is done by an extension or supplementation of the evidence, which remains in accord with the evidence (or claims to do so), but carries it further by utilizing other sources of information besides those which started the whole process going, and, in many instances, in addition to those which can be directly identified in the external surroundings. Between the initial information and the terminal stage, when the gaps are alleged to be filled, or completeness achieved, theoretically there is always a succession of interconnected steps...they, more than anything else, are what makes an experimental approach to thinking theoretically possible, and they confer upon thinking its character of necessity' (p. 75).

Now, like most working definitions, this one is not a particularly felicitous one. For as C. S. Peirce so wisely insisted long ago, and Wittgenstein more recently, the definition of a thing or process is the array of predicates one can launch in the sentence frame 'if x, then...'. But it is adequate to underline what Prof. Bartlett conceives to be formally central: that there be evidence with gaps and that these gaps be filled in a manner which, to the thinker, is 'fitting' or 'adequate'. The gaps exist, we may infer, by virtue of 'direction'. Without direction there would be such an entropic hodge-podge that one could not tell the gap from the chaos. Which leads me to my first major criticism of this provocative work. It is that too often (not always), the author adopts a kind of Baconian or common-sense view to the effect that direction (or what I would prefer to call 'structure') somehow grows out of evidence as the autochthonous gods were supposed to grow out of the hallowed soil of Olympus. At his most negligent he speaks merely of the 'extraction of rules' from evidence, without examining in sufficient detail what is involved in such extraction.

What are the great principles of transforming informational input, taming it into a structure, imposing direction upon it? I believe that the 'skill' paradigm to which Sir Frederic has wedded this enterprise keeps him from staring boldly enough at this problem. When he returns to the familiar hunting ground of memory organization, the sparks fly. 'All the way through this book I have taken the view that thinking is possible only if much that precedes it in an order of development continues to exercise an influence. This is true in that the thinker is able to avail himself of evidence no longer a part of his immediate external environment, and in that the ways in which he treats the evidence very often have conventional characters which could not possibly be constructed on the spur of the moment. Perhaps the very most critical of thinking problems have to do with what sort of processes of storage of evidence thinking demands, and what are the implications of the uses thinking can make of the stored evidence' (p. 87). Here would have been the staging area for a revisit to the *schemata* which stood Sir Frederic in such good stead in his classic volume twenty-five years ago. As it is, he has some most arresting things to say about the grasping of similarities in contrast to differences for the construction of viable thought structures (a term he does not use), viable meaning 'transferable'. See, for example, one specific difficulty introduced by the 'skill model'. Sir Frederic notes that, given direction in thinking, there is a point of no return as in skilled movement. This is indeed the case. But it is also the case that the models that we use in locating direction have as one of their most interesting properties the character of reversibility. Two plus two equals four, but four is also decomposable into these two factors. 'Some p's are q' has the reverse implication that 'some or all q's are p'. The structures that underlie thinking have neither a property of timing nor of 'no return'. It is only when we base an hypothesis upon them that these properties emerge. As I shall urge again before ending this review, the properties of these underlying structures or schemata are much in need of study if we are to understand the thinking acts that are based upon them.

The book deals with four types of thinking, each of which may be conceived to have a somewhat different objective and each a recognizable strategy. The first is 'closed-system thinking'. In rough terms, closed-system thinking involves the attempted solution of well-defined problems of a formal or highly structured nature, and such attempted solutions involve either the interpolation of information in a structure, as in supplying the numbers between 2, 4, 8 and 128, 256, 512 or in extrapolating from a series such as 1, 3, 5, 7, 9, 11, 13,.... Closed-system thinking may also involve transforming 'evidence in disguise', which consists of recognizing the relevance of information to a problem although the information is organized in a fashion not immediately relevant to the problem. Or it may mean that the information must be manipulated in order for a problem to be recognized.

Prof. Bartlett's demonstrational experiments in this portion of the book, fresh and ingenious though

they be, leave one a bit puzzled. They show beyond question that interpolation and extrapolation occur, and that subjects use a highly limited array of solutions to such problems. Why? Surely it is not for Kantian reasons, that there are only a limited number of categories of thinking. I could not help but ask myself what it was that we knew about the *acquisition* of the solution models that people apply to closed-system problems. Very little indeed. Perhaps the answers must come from studies of the learning or acquisition of 'rules' or 'logical models'. One misses reference to such work, particularly the researches of Piaget.

A second type of thinking is more 'adventurous' and is called *experimental*. 'His [the thinker's] aim is not simply to extend and define knowledge within a system that is to be treated as already completed, but to help to develop and understand systems which are so far literally incomplete' (p. 98). The discussion of experimental thinking is, in the reviewer's opinion, the wisest and most penetrating set of reflections on the subject that has ever been put to print—flexible, informed by much direct experience as well as by some interesting experiment, full of challenging notions about the use of past experience in transfer. Much emphasis is placed on strategies for spotting the similarities between events as a basis for discovering common features in seemingly disparate problems. For it is the discovery of such common features that makes transfer of solution possible. Here I think Sir Frederic has pointed his finger directly at the reason why in thinking and conceptualizing, people prefer to work with 'positive instances', examples that illustrate the principle they are working on. For these permit us to transfer what we have already learned, whereas negative instances leave us with nothing to transfer, force us back into a kind of trial and error. As James Bryant Conant once remarked, the object of science is to reduce empiricism—to allow us to spot and exploit the similarity between that which we encounter anew and what we already know.

The third type of thinking discussed, 'everyday thinking', has as its aim something else, something of central import to the social psychologist. 'The closed-system thinker and the experimental scientist are set to achieve results which carry with them an internal compulsion. Everyday thinking also aims to arrive at a compelling issue, and, however it makes its journey, to stop where every other normal person must stop. But the steps to the resting place are never all of them exhibited, and there is no inherent compulsion in their nature and relations. It is not in the nature of the assertion, but in the manner of assertiveness that everyday thinking seeks to attain necessity... the source of the compulsion being now within the thinker, and particularly in his social group, it is possible, and indeed common, for completely contradictory issues to claim the same necessity.' Sir Frederic then goes on to say that everyday thinking might as appropriately be called 'immediate communication thinking'. 'Immediate, here, means dealing with highly topical events....The communications are then likely to share the urgency of the events, and thus to favour a strong and unconditional mode of expression. The same mode is still further strengthened by all those self-regarding sentiments from which nobody who enters into the "immediate communication" situation can escape.' In achieving generalization, even in the absence of social convention, it proceeds not by a 'painstaking, slow, and intelligent comparison of many instances, and the abstraction of common features, but by an easy and natural exaggeration which in the extreme may be based upon one single impressive personal experience... where formal or experimental structures are concerned..., the few instances are examined for whatever "rules" they may reveal. In everyday thinking the instance, or the few instances, are not so much explored as, so to speak, "savored" and used. The generalizations of the non-exploratory kind have little to do with transfer of practice or training save that they make it more difficult.'

The discussion and observations on everyday thinking are highly challenging. My only complaint is that Prof. Bartlett makes social communication seem like such a *barrier* to adequate thinking. As Arieti has recently argued (and Piaget, Vigotsky and George Herbert Mead before), it is the need for adequate social communication that leads people to learn and follow the rules of logic and inference as well as conforming to the primitive principles of denotation. Finally, Prof. Bartlett would have found a powerful ally for his contentions about the conventionalizing effects of social communication in Benjamin Lee Whorf whose work on psycholinguistics shows how deeply the socialized forms of assertiveness affect the texture of thinking.

Let me finally revert to several general points that are put in issue by this challenging and pioneering work. The first has to do with the manner in which the process of thinking itself may be most fruitfully formulated. As Sir Frederic insists, thinking involves the use of past experience, and I would add *organized* past experience. We organize our past experience in such a way as to get 'maximum travel' from it. Such organization results in our being able to subsume new experience under old rubrics so that we need not deal with the new experience *de novo*. We convert past learning into something generic,

and having grasped that $A > B$ and $B > C$ implies that $A > C$, it makes no difference whether we are dealing with ergs, potatoes, or Irish whisky. To be sure, the structures or schemata used in thinking are altered by the range of instances to which they are applied, but what is basic to thinking is that such underlying structures or 'rules' or 'transforms' are necessary.

The process of thinking, then, may be conceived in two steps: the one is a process of rule or structure learning, the second is a process whereby we recognize events as conforming to the structures we have learned. It is about the former process that Prof. Bartlett writes in discussing 'adventurous thinking', about the latter in connexion with closed-system thinking. Yet, one gets the impression that there is some confusion between the two that leads him to speak of extracting 'rules from evidence' in discussing the way subjects deal, for example, with well-defined arithmetic problems.

I would urge that at least two missions should be marked out for the psychologist concerned with thinking. One is to learn how organisms acquire—invent, discover, learn, mature into—the structures or schemata that we apply in thinking and problem solving, The genetic approach is one fruitful avenue, most recently attested to by *De la logique de l'enfant à la logique de l'adolescent* of Inhelder and Piaget. A second approach is phylogenetic, and I would cite Klüver's *Behavior Mechanisms in Monkeys* as a stunning example. Prof. Bartlett's account of adventurous or experimental thinking is yet another approach—a close analysis of how people operate in their thinking when they are at the frontiers of knowledge. His exciting account of the development of *Remembering* is a prime document.

The second mission is to examine the strategies whereby people are enabled to recognize events as exemplifying the principles or structures they have learned so that these principles or structures may be applied. How, in short, do we achieve what William James once called 'the electric sense of analogy'? How do we avoid seduction by the irrelevant, confusion by the seemingly disordered, etc? Sir Frederic's chapters on thinking in closed-systems suggest some interesting answers to these questions.

In fine, *Thinking* is a book whose appearance gives promise that the experimental psychology of thinking is emerging as a vigorous field of inquiry. It is a bold overture, one that will have an impact on future research.

PUBLICATIONS RECENTLY RECEIVED

Schools of Psychoanalytic Thought. By Ruth L. Munroe. London: Hutchinson. 1957. Pp. xvi+670. 42*s*.

The authoress of this book is a clinical psychologist in New York, specializing in diagnostic tests. She tells us that she has collaborated professionally with 'some fifty' psychoanalysts 'of various schools'; that she is the wife of Dr Bela Mittlemann, an established analyst, with the marriage perquisite of being able to listen-in to shop talk; and that 'as part of my training as clinical psychologist as well as for personal reasons, I have been in analysis with a fairly classical Freudian and at another period with one of Horney's first lieutenants'. Her chief purpose in writing this book seems to have been to evaluate and compare those analysts 'whom most of us in America think of as having established schools'. To do this, she has carefully and strenuously attempted to present 'an objective account' of these different schools. 'In writing', she tells us, 'I have tried to feel myself into each system as a whole, thinking for the nonce in the terms of its major premises and working through to its conclusions as if I were successively a Freudian, an Adlerian...' and so on. In her attempt to achieve expository accuracy and fairmindedness, she took the step of submitting her draft account of each school 'to experts selected for their special acquaintance with the school'. When she comes to evaluate and compare, she adopts 'a critical stance' and attempts an 'integration' of the schools she has presented. In doing this, she passes from exposition to clearly separated comment. This comment, she says, is of 'two orders, often closely commingled'. First, she presents the reflexions of a clinical psychologist about a school in order 'to invite thought and inquiry in areas which, I feel, have been neglected by the school'. And secondly, she attempts 'a formalistic rephrasing of the basic problems of psychological science' as a means of 'integrating the schools reported' in this book. If I understand her rightly, this rephrasing is part of an attempt to develop a more inclusive theory or perspective than the theories of the schools that she reports. For this purpose, she has 'borrowed many ideas from contemporary philosophers, notably from the romanticists and the logical positivists'.

The plan of the book is as follows. She begins with a short introduction in which she explains the general nature of the enterprise. In the next chapter, she extracts and explains the concepts that she thinks basic to and accepted by *all* brands of psychoanalysis. These are four in number: psychological determinism, the role of the Unconscious, the goal-directed character of human behaviour, and the genetic approach to personality. She then goes on to consider the school of 'Freud and the "Freudians"'; and this makes up the largest single section of the book. By 'Freudians' she means those analysts who accept the 'libido theory'. She then examines the 'non-libido schools' of Adler, Horney, Fromm and Sullivan. Her main concern is with the cleavage in the United States between 'libidinal' and 'non-libidinal' analysts. She continues with a short examination of Jung and Rank, who 'do not fit this libido-non-libido dichotomy'. She ends with an 'Epilogue' in which she puts together her suggestions for integrating the psychoanalytic theories she has examined.

I shall concentrate on the chief concern of the book. What are the main comments she makes on the libido and non-libido schools? What integrating 'rephrasing' does she offer? I pick out three points and themes on which she criticizes Freud and the 'Freudians'. (*a*) This theory, even in its current development, overlooks or pays insufficient attention to 'drive systems' that are non-sexual in character. By a 'drive system', Munroe appears to refer to a systematic pattern of goal-directed activity, closely connected with some bodily apparatus whose functioning 'involves inner tensions', and which are 'oriented toward the external world' for the completion of the systematic pattern of activity. Obvious examples are the respiratory, cardiac and digestive systems of the child; less obvious examples, perhaps, are its neuro-muscular (or 'motility') and speech systems. In overlooking these other non-sexual systems, Freudians are apt to forget (in theory if not in practice) that 'from the very beginning the infant is dandled, talked to and stimulated visually, usually by the adults who also care for its "sexual" systems'. Because of this, Freudian theory misses much in the relationship between child and adult that may be, and probably is, important. Munroe suggests, for example, that a good relationship between them and, consequently, a favourable development in the non-sexual systems 'will suffice in the majority of cases for the child's lonely handling of crises in the more global "sexual" systems as they arise in the course of libidinal development'. In its exclusive emphasis on the erogenous zones, Freudian theory is guilty

of a major error in Munroe's eyes. This is the error of 'reductionism'. To commit it is 'to apply a perfectly sound idea too generally, with neglect or exclusion of concomitantly operating systems of events'. Scientific theorizing, she argues, should be multi-dimensional in character. It should study data 'along many lines (dimensions), which are then related to one another'; and thereby avoid the type of integration that 'reduces "everything" to a master principle'. (*b*) Early Freudian theory is incomplete in not giving enough recognition to the importance of 'ego' functioning. She discusses Hartmann's suggestions with approval. She maintains that there is 'a major omission from Freud's statement of the important dynamic groupings of systems (structures, institutions) that come to function more or less independently in the mature personality. This is the self image', and she explains the relationship between this concept and the super-ego and the (neglected) ego-ideal of Freudian theory. (*c*) In discussing pathology, she declares that 'for every case that neatly fits the simple Freudian dynamics here outlined, there are ten for which the interpretation requires drastic modification.... Any simple extrapolation from libido theory to the majority of cases encountered in unselected practice comes to grief very quickly.'

Her comments on the non-libido schools complement those she makes on the Freudian. Consider Horney alone. She approves of Horney's concept of the 'idealized image of the self'. This and what goes with it encourages the development of 'ego-psychology', with its emphasis on the interpersonal relations of the here and now; and she remarks that Freudians 'are just beginning to give proper theoretical weight to concepts which Horney has been shouting from the house tops for twenty years'. But she criticizes Horney on two counts among others. (*a*) She commits the error of reductionism in making 'the need for security' the basic principle by reference to which all the variations in personality development are elucidated. (*b*) She is quite wrong in rejecting the Freudian thesis that the sexual drive systems have an important role *per se* in personality development and later dynamics. (The same criticism applies, *mutatis mutandis*, to the other theorists of the libidinal school.) 'There is every reason', Munroe writes, 'to suppose that the biologically determined sexual systems *do* have an autonomous course of development which must profoundly affect the relationship of the young child to his environment during the early months and years of his life—and therefore profoundly affect the attitudes he develops along the lines emphasized by Horney herself.' In view of Munroe's own emphasis on the importance of the non-sexual systems, she goes out of her way to declare that ' I cannot state too emphatically my judgment that the infant *is* overwhelmingly concerned with its profound biological functions (oral, anal and genital); that its early interpretation of the world *must* be primarily constellated round these functions'. This dual attitude enables her to say that 'there are probably relatively few cases (i.e. children) in which the sexual systems *as such* actually provide the overwhelming focus for the personality as a whole in the manner suggested by the earlier Freudians'; *but* that ' there are also relatively few cases in which the developmental systems that these schools emphasize so thoroughly determine the personality as to make the course of the sexual systems in themselves a matter of indifference'.

These comments by Munroe help us to understand the sort of 'integration' she espouses. She urges on us a more inclusive 'perspective' in which emphasis is laid on the organismic self (or total personality) operating through different sorts of 'systems'—the 'drive systems' of the organism, the 'dynamic sub-systems of very great variety' which develop in the course of living, and the 'transient sub-systems' established moment by moment. This organismic self develops 'a self image' which has 'an underlying dynamics' of its own in our culture that merits study. Our drive systems are sexual and non-sexual in character, and the latter also have 'an autonomous dynamics' that requires further investigation. What is essential is that we avoid premature universalization and tight, rationalistic theorizing. Her more inclusive perspective and rephrasing can be best appreciated, perhaps, by noting how she rejects the one-sidedness of both the Freudian and non-libido schools, and how she attempts to synthesize their constructive contributions. She ends by extending her 'view of systems' to include the difficult problem of aggression—on which she has interesting suggestions to offer.

My brief sketch of her argument does not begin to do justice to the mass of comment and suggestion the book contains, or to the 'integrative rephrasing' she offers. I have said nothing, for instance, about her approval of Kardiner and Anna Freud, of her criticisms of Melanie Klein, and of her remarks on Sullivan (whose line she considers the most promising of the non-libido group), and of her recognition that Jung's *persona* is close to her concept of self-image. In this review I have only been able to pick on one central contribution that the book offers us. It is fair, I think, to warn the reader that, in her endeavours to be fairminded, Munroe has been longwinded and a bit repetitious—the book has the laboured air of careful toil at midnight.

What are we to make of the central contribution I have outlined? Let us ask the preliminary question: 'Whom is she addressing?' In so far as she is writing for young psychiatric registrars, for students of

psychiatry and clinical psychology, for the younger psychoanalysts, and the like, she will probably not be unsuccessful in inducing them to be more tolerant of alternative points of view in psychoanalysis, and to appreciate better the contributions that each has to make. Hence the book may do a little to bring about a rapprochement between the practitioners of the various schools. But in so far as she is writing for established analysts, for people who have been through years of training (or 'indoctrination', as some allege) in some School or Institute of psychoanalysis, I doubt whether the impact of the book will be great. Likewise, in so far as she is addressing the hardheaded, empirically minded layman, or scientific worker, I doubt whether the impact will be at all considerable.

I do not think the book will move the established analyst for the following reason. When I ask myself: 'What grounds does she offer for her integrative line?', I find it difficult to discover any. I suspect that the grounds that move Munroe to talk as she does are: (1) that this line synthesizes her own clinical experience and the alternative psychoanalytic theories in a way she finds personally satisfying; (2) a hunch that her line will help the development of psychoanalysis at this juncture. If I am right about this, then it is evident that the whole book has the strength and weakness of the usual story produced by a clinical psychologist or psychiatrist about a patient or a set of clinical data. The story may be very suggestive and illuminating, and it may even cover all the known facts coherently. But it is not very compelling, especially to established practitioners with developed and pet stories of their own that they prefer to spin. What *would*, I think, move them would be to show them that some important thesis of theirs is *false*. I take it that she wants to say, for example, that the Freudian thesis about the exclusive role of the sexual drive systems is a false thesis. But does she *show* this? I hesitate to say so. What is rather futile is for Munroe to accuse the Freudians or Horney, or anyone else of 'reductionism' when she does not provide an independent criterion to decide whether an analyst is 'reducing' or not. In the absence of such a criterion, the word 'reductionism' functions, largely though not wholly, as Munroe's favourite smear word.

The book is not likely to move the empirically hardboiled for similar reasons. They are only too likely to ask: 'Why all these different schools of psychoanalysis? Why all the disputation? Especially as Munroe speaks about the techniques of the psychoanalyst as if they were quite satisfactory?' It is not enough to answer (as she might): 'Because psychological science must work with loose theories or systems, unlike physics.' For what, precisely, makes a system a 'loose' one in science? It looks as if the theories, or systems, of psychoanalysts are *so* loose that it is not possible in principle to take any steps to decide between them, in which case they cease to be loose systems *in science* and become systems of a literary and metaphysical type. But she would repudiate this contention. She wants to talk about a psychoanalytic theory 'as a provisional systematisation of events', rather like one of the theories of light, which means that it makes some claim to objective adequacy. But then she has to meet the challenge of verification—how does one set about confirming or disconfirming any one of these theories? She appears by implication to hold that this can be done by means of the data of the analytic session and the projective test, and the data from the direct observation of children. She makes constant and (to me) inconclusive references to data of this sort. Thus, she seems to imply that this sort of data can establish that motility, and other non-sexual systems, play the role she alleges they do; and that the infant's early interpretation of the world is constellated round its oral, anal and genital functions. But to take this line is to plunge into all the complexities of psychoanalytic method and all the uncertainties involved in interpreting the behaviour of children. Munroe does not deal with these complexities and uncertainties. So the book is likely to strike the hardheaded as theoretically and empirically shoddy. I think the book raises very difficult issues about the whole character of psychoanalysis and psychoanalytic theories. These are issues of logic, and I cannot help feeling that they are beyond the range of Munroe's competence.

But I have carped enough. I end with a note of congratulation. In writing this book Munroe tried to think herself into the various theoretical positions in psychoanalysis. I think she deserves to be congratulated on this aspect of her enterprise, and on the success she has achieved. To think oneself into the various theories in some domain is not an easy task, especially where some of the theories are personally repugnant. But it is an excellent discipline. I recommend it to any psychoanalyst and psychologist of personality, and especially to those of the combative type.

B. A. FARRELL

Backwardness in Reading. By M. D. VERNON. London: Cambridge University Press. 1957. Pp. viii + 228. 25*s*.

In her latest book, Prof. M. D. Vernon has produced a scholarly, fully documented and impartial appraisal of the present-day state of our knowledge about children's backwardness in reading. Although she deals with many technical researches, she has set out the material in a clear way so that the book

is not only of help to educational psychologists, but also to classroom teachers. The latter may lack some of the knowledge required in connexion with some of the more physiological and psychological aspects of the problem, but will, none the less, be greatly helped by Dr Vernon's clear statements and her full discussion of the educational issues involved.

Dr Vernon takes us first through research on the perceptual basis of reading and these chapters could well be read by psychologists for useful evidence relating to perception. She then discusses the nature of reading disability, emphasizing the fundamental difficulty experienced by backward readers in connecting up visual perception with auditory perception and with the cognitive generalizations and meanings it is hoped to develop. In discussing the innate factors underlying disability in reading, it is heartening to find Dr Vernon writing so firmly about the limitations of the blanket term 'congenital word blindness'. The term really conceals ignorance and the less it is resorted to, the healthier our research into the problem of reading difficulty will be. She then takes up the effect of various acquired difficulties in reading ability. This is an informative chapter and one can question her views only in one small respect: this is on page 130 where she hazards a guess at the effect of premature birth upon linguistic and visual functions. A recent large-scale survey in Birmingham of 7 to 8-year olds all born prematurely, *excluding* all cases of brain damage and other abnormality, revealed no significant difference in scholastic attainment between the premature sample and the control. In the last two chapters, Dr Vernon discusses environmental factors springing from the home and school and also the problem of the remedial treatment of reading disability. In the case of the problem of remedial education, she presents the material in a detached and well-balanced way. She wisely eschews a discussion of the more controversial matter of the best measure of backwardness.

Dr Vernon provides the reader with a critical conclusion at the end of each section and then sums up in a last section on the over-all position, and she squarely faces up to the fact that methods of experimental and clinical study have now 'failed to isolate any factors which appear universally in all cases of reading backwardness—other than inability to read easily!' Later we are told that 'the only fact which appears clearly at first sight is the heterogeneity of cases of reading disability—heterogeneous both in the origin and in the nature of their disability'. Causation is multiple and perhaps the only general conclusion to be made is that reading demands intellectual capacity and that disability in reading is shown by 'the child's general state of doubt and confusion as to the relationship between the printed shapes of words, their sounds and their meanings'. Dr Vernon also emphasizes the highly important influence of motivation in reading attainment, pointing out that maladjustment and reading failure go together, but that the evidence is insufficient to show how far one is caused by the other. This book is to be recommended as a comprehensive assessment of our present knowledge about reading disability. Prof. Vernon has provided us with yet another scholarly analysis of an important field in psychology and education.

E. A. PEEL

The Slow Learner: Some Educational Principles and Policies. By M. F. Cleugh. London: Methuen. 1957. Pp. vii+186. 13*s.* 6*d.*

The subtitle 'some educational principles and policies' is less misleading about the contents than the main title. The book is not about method, and is especially intended for administrators, who, according to the blurb, 'have the responsibility of organizing the provisions for special educational treatment'. As a result, Miss Cleugh ignores the basic problem of organization, which hinges upon the responsibility of educational psychologists in these areas. Even when a plan for systematic ascertainment is reported and commended by the author, the fact that it was carried out by a psychologist is mentioned only incidentally. Similarly, the book contains many examples of misclassification by medical officers and yet the obvious first rectifying step of making psychologists at least informally responsible for intelligence testing is not suggested.

Miss Cleugh's intentions are admirable and she successfully simplifies some of the complexities, while emphasizing the varied possibilities of helping dull and retarded children. The book can therefore be recommended as an introductory text for those students contemplating a career in educational psychology, although some of its language and ideas are rather out of date. In a discussion of I.Q.'s, children are described as having 'half the normal intelligence', and this may well encourage incorrect conceptions of test results. Miss Cleugh also takes it for granted that all mongols are ineducable, whereas most practising psychologists will have met exceptions to this rule. When it comes to emotional reasons for retardation, the author subscribes to the idea that it is pointless to deal with a symptom while neglecting

a cause. This psychiatric creed is open to experiment and the results might have important effects on the future development of child guidance. In the opinion of many teachers the best thing to do with children retarded in reading is to teach them reading, even though the original cause of retardation was quarrelling parents, a traumatic experience or what-have-you. In fact, one might go further and say that remedying the result might affect the cause, for it would not be surprising to find parents quarrelling less after their children had shown an unexpected academic success.

DAVID SCHONFIELD

The Dynamics of Anxiety and Hysteria: an Experimental Application of Modern Learning Theory to Psychiatry. By H. J. EYSENCK. London: Routledge and Kegan Paul. Pp. xiv+311. 1957. 32*s.*

Two very interesting trends in the recent history of psychology are illustrated by this book. The first is the growing feeling that it is high time for a *rapprochement* between the sort of psychology that looks for general laws governing all human (or mammalian or animal) behaviour and the sort that studies individual differences. Hitherto they have been pursued by distinct groups of psychologists, who have not kept abreast of each other's work and who have used such vastly different languages and methods that one would scarcely believe both lines of research to be directed at the same organisms. Eysenck intends the present work as a contribution to the removal of this anomaly. It represents, he says, a transition from 'classification to a study of dynamics...from nosology to aetiology, from description to causation', as far as his own development is concerned.

The second trend is the spread of a changed view of the relations between learning theory and psychoanalysis. At first, learning theorists were impressed with the common ground between their conclusions and those of Freud, despite the very different kinds of data from which the two were derived. They produced theories of human behaviour, methods of psychotherapy and accounts of what happens in psychotherapy which, while they embodied many divergences from Freud's matter and especially from his phraseology, preserved much of the picture that he had painted. The group associated with Hull at Yale were particularly diligent in this regard, but Tolman, Skinner and Guthrie also contributed. During the last few years, however, other learning theorists have been making vehemently unfavourable remarks about psychoanalysis and advocating techniques of symptomatic therapy that flatly contravene most of the rules that have been widely accepted ever since Freud laid them down. The way in which Dollard, Miller and Mowrer, on the one hand, and Eysenck, Taylor and Wolpe, on the other, can derive such diametrically opposite theories of abnormal behaviour from the same Hullian type of learning theory is highly instructive. It suggests that rigorous deduction must be distinguished from plausible inference and that there is no substitute for the patient accumulation of empirical data under controlled conditions. It suggests also that some 'halo' and 'horn' effects have been at work. Freud said so many things that it is highly improbable *a priori* that all of them are true or that all of them are false. What seems to have happened is that learning theorists have picked on certain elements of what psychoanalysts say, found these elements acceptable or unacceptable, and consequently concluded that the bulk of psychoanalytic theory must be on the right or, as the case may be, the wrong track.

The title of this book of Eysenck's must not be taken too seriously. This is a pity, since many criticisms will probably stem not from the intrinsic value of the contents but from the way in which the contents disappoint the broad expectations that the title invites. The book is really volume v in a series of periodical summaries of research carried out at the Maudsley Hospital's Psychological Department. The latest phase of this research has concentrated on measurable differences between introverts and extroverts, including sufferers from anxiety states and from hysteria respectively. It has been inspired by aspects of Pavlov's progressively changing theory of personality types and by Hull's theory of inhibition. But it does not make any contact with the work of Pavlov and others on experimental neuroses, with the voluminous studies of particular human syndromes reported in Pavlov's *Conditioned Reflexes and Psychiatry* or with the substantial body of experimental work on anxiety as an acquired drive, on conflicts and on reactions to frustration that has grown out of Hull's ideas. It does not tell us much about the reasons why anxiety neurotics and hysterics differ from other introverts and extroverts, and it offers only a few perfunctory, *ad hoc* and admittedly speculative suggestions respecting the reasons why some individuals have one set of symptoms and others have a different set.

Not only is there no attempt to cover all aspects of the field denoted by the title, but the policy of using the book to report all the Department's major research projects that have not been written up in previous books produces a not-too-satisfactory compromise between a unified and systematic coverage of one topic

and a series of relatively independent papers on different topics. There is a tendency to dwell in excessive detail on matters that are not directly relevant to the main theme of the book. They may be directly relevant to matters that are directly relevant to the main theme. But that is not the same thing.

The crux of the book lies in the hypothesis that excitation is generated more rapidly and inhibition more slowly in introverts than in extroverts. It was first put forward by Eysenck some years ago with the support of Franks's experimental findings, which showed that eye-blink conditioning occurred more rapidly and extinguished more slowly in introverts than in extroverts. The reaction in many quarters was one of scepticism. For one thing, it was felt that the experiments should be repeated with other conditioned responses, in order to verify that the findings are not confined to the blinking response and that there is such a trait as general conditionability. Secondly, there are many factors that could produce differences between individuals in speed of conditioning and of extinction. They include, besides the rate of growth and asymptote of inhibitory potential (which Eysenck favours), drive level, amount of reinforcement and stimulus–intensity dynamism. Variations in any of these might conceivably account for the observed differences between introverts and extroverts. Moreover, a demonstration that introverts and extroverts differ in one of these factors would not exclude the possibility of their differing in one or more of the others also. The main distinguishing characteristics of inhibitory potential are that it accumulates with prolonged evocation of a response and that it dissipates in the course of a rest period. Many readers of Eysenck's original papers on the topic wished, therefore, that data comparing reminiscence after, say, a 24-hour interval had been secured from Franks's subjects.

The hypothesis that a higher level of anxiety drive is what makes introverts acquire eye-blink conditioning more readily is in line with what Spence and his associates have been writing. It is a tempting hypothesis in view of the concentration of both Franks and Spence on noxious rather than rewarding unconditioned stimuli and of the fact that neurotic introverts are characterized by 'manifest anxiety' symptoms. Eysenck criticized Spence's position, and he reiterates this criticism in the present book, on the grounds that some neurotics, i.e. hysterics, show slower-than-average conditioning. But this would contradict Spence's hypothesis only if it could be shown that neurotics, including hysterics, have a higher drive level than non-neurotics, a speculation that Eysenck has tentatively offered from time to time but has never actually demonstrated.

The two pieces of research that would answer the most persistent sceptics have still not been carried out. But in the meantime, Eysenck has marshalled a most imposing collection of findings, from such varied fields as perception, decrements in prolonged motor performance, mental fatigue and drugs, all of which point in the direction of a greater proneness to inhibitory processes in extroverts than in introverts. He freely admits that each of these studies by itself may be inconclusive and points out flaws in many of them. But the net effect of all these data is to win a serious hearing for his hypothesis that inhibition holds the key to the introversion-extroversion continuum.

Throughout the book, Eysenck submits his conclusions with considerable modesty and humility. He is generous to those whose work he is attacking, except for the psychoanalysts. He reiterates that he does not regard his theory of individual differences as established. He merely claims that it will yield enough testable predictions to be scientifically respectable and declares his willingness to abide by the verdict of future research. It must be admitted that he anticipates and acknowledges most of the criticisms to which the material in his book may be open.

In short, this is an important book that all experimental psychologists, especially if they are interested in learning theory, all personality theorists, all clinical psychologists and all psychiatrists ought to read. It reports many experimental results and theoretical innovations that are of interest in themselves, quite apart from their bearing on the author's theory of individual differences. The ideas it introduces certainly deserve to be thrown into the arena to hobnob with those that have hogged the limelight hitherto.

D. E. BERLYNE

An Introduction to Psychopathology. By D. Russell Davis. London: Oxford University Press. 1957. Pp. vii+388. 30*s*.

This is an attempt to bring together the disciplines of experimental psychology and clinical psychiatry. The author's training fits him well for this task: he has spent much of his time in the Psychological Laboratory at Cambridge while holding a Consultant appointment at the Fulbourn Hospital. A synthesis of this sort is always an ambitious undertaking: Masserman, Hunt, Dollard and Miller, Mowrer and others have all made similar ventures with varying success. It cannot be said that Dr Davis has been any more

successful than those just mentioned, for he has imposed on himself certain strategic limitations that give rise to an unduly rigid and doctrinaire treatment of the subject. In the first place he has entitled his book 'Psychopathology', thereby committing himself to a one-sided approach to mental illness and mental disorder which forces him to minimize the role of physical and genetic factors to an absurd degree. Furthermore, in giving a rather special meaning to the term psychopathology, he adopts an unduly mentalistic attitude both as regards the phenomenology and the aetiology of the various states described, besides perpetuating a physical-mental dichotomy that is surely by now out-dated. But Dr Davis adds to his difficulties by other defects. He has made little effort to organize his material systematically for the benefit of his readers. At times, he appears to address the psychiatrist, uninformed in the aims and methods of the experimental psychologist; at others, he addresses the experimentalist in like fashion and, at others again, he writes for the student or intelligent layman. There is nothing in common with these various essays except that they are written by one man. They exemplify the more unfortunate consequences of too loosely committed eclecticism. Then there are a number of irrelevancies appended for no very good reason beyond the fact that they constitute a particular interest for the writer himself.

Dr Davis's main thesis is that 'psychopathology is dependent on psychology and draws from psychology the principles by which to explain the diverse and unsystematic observations made during the investigation, treatment and follow-up of patients. These principles, established and proved in the laboratory, are applied in the clinic' (p. 2). He sets out to develop this theme in two parts. In Part I, he deals with various aetiological factors in mental illness and mental defect, stressing the importance of family structure, early environment and cultural influences. He gives an account of mental development and its various disorders. Thus this part is largely a discussion of clinical entities, loosely arrayed, with special reference to the influence of 'environment' conceived along fairly conventional lines. In Part II, the author describes past and current experimental methods having to do with behaviour disorders. Invoking neo-behaviourist researches and theories, he sets out to explain mental disorders in terms of these theories. In the process, both the impeccability of the experimental procedures and the richness of the clinical material suffer a sea-change so that neither universe of observation is improved. As an experimentalist the author falls into his own trap of woolliness and as a clinician retreats into his ivory tower of objectivism. Here and there throughout the work the author alludes to the contributions of psychoanalysis. He concedes that most psychopathologists owe more to Freud than they are apt to admit. He himself does not owe much; nor does he appear to have grasped the essentials of Freud's discoveries, notably with regard to the unconscious aspect of psychological processes. It is indeed surprising to encounter (p. 32) in 1957 an objection to psychoanalysis that was abundantly refuted by Freud in 1914: that Freud's original observations were made on upper-class Viennese with sexual conflicts more overt than in present-day England! As is now so commonly found in works of this kind, the section devoted to therapy is short and quite unhelpful, confined as it is to intellectual acceptance on the part of the patient and moral exhortation by the therapist armed with a novel and intricate terminology that should impress.

It must be recognized that the difficulties inherent in a synthesizing approach to psychology and psychiatry are immense and that it will be many years before we can attain to a truly synoptic account of these two disciplines. The author must be admired for his heroic attempt to bring together in one volume such a disparate collection of facts and theories, but reluctantly it must be said that the disparity remains and confusion is increased.

W. M. MILLAR

Group Psychotherapy: the Psycho-analytical Approach. By S. H. Foulkes and E. J. Anthony. Harmondsworth: Penguin. 1957. Pp. 263. 3*s.* 6*d.*

Though group psychotherapy is coming to be regarded as the most promising development during recent years in the treatment of psychiatric disorders, and is being increasingly adopted in hospitals and clinics, descriptions of its theory and practice have, in this country, been mainly confined to individual papers scattered through the journals, and no clear exposition for the lay reader has existed. The authors have undertaken the difficult tasks of attempting to bring together an outline of current theory and technique in the field for the information of specialists, while at the same time making their exposition simple and clear enough to enable the educated layman to follow most of it without much previous acquaintance with psychiatric theory and jargon. The compromise has been achieved skilfully, and, though it makes for a certain unevenness, it is probable that the book has gained rather than lost in value through this need for clarity in expression. Both authors are peculiarly suited to the task they have set themselves since Dr Foulkes was the first to develop analytically orientated group therapy in this country, while

Dr Anthony has extended its original use with adults to disturbed children and adolescents, and for purposes of research into the nature and origins of mental disorder. Particularly valuable is a lengthy 'introductory survey' which traces the historical development of group therapy and the debts it owes to psychoanalysis and to various earlier studies of social interaction, such as those of Kurt Lewin, Trigant, Burrow, and Lawrence Frank. It also contrasts the theory and technique of Slavson, Wolf, the Tavistock Group and other workers with that evolved by the authors and termed 'group-analytic psychotherapy'. The discussion is admirably fair, detached and balanced, a welcome change in psychiatric fields, where workers are often prone to argument based on personal rivalries and animosities. It emerges clearly that essentially similar theories and methods have evolved from the various 'schools' of group therapy, the main difference lying in the degree to which treatment of individuals *in* a group by the doctor has evolved into treatment *of* the group *by* the group, the doctor becoming increasingly passive and interfering only when the process is blocked, or when his special knowledge is required to help them translate and understand the answers they have produced unwittingly. The authors claim to have progressed furthest in this direction, and operate on the assumption that '...the group tends to speak and react to a common theme as if it were a living entity, expressing itself in different ways through various mouths. All contributions are variations on this single theme, even though the group is not consciously aware that this is so and do not know what they are really talking about.'

The remainder of the book is devoted to the changes in therapy that this revolutionary discovery effects, and to detailed descriptions of technique in its practical application. Most of this is clearly written and full of valuable information even for those with much experience of group psychotherapy, though some of the theoretical speculation, particularly the attempt to describe the process in Gestalt 'figure-ground' terms, introduces unnecessary complication without really explaining anything. The weakest sections are the clinical illustrations. The verbatim records of children's groups are not accompanied by adequate explanations of the latent meaning of what was said, or the purpose it served, while the long description by a member of an adult group of her experience in it is unconvincing to the inexperienced reader, and would have been better replaced, or at least supplemented, by extracts from tape recordings which are readily available and can show the various phenomena encountered in groups with convincing clarity.

R. SKYNNER

Family and Social Network. By Elizabeth Bott. London: Tavistock Publications. 1957. Pp. xi+252. 30*s*.

This is an account of an investigation conducted by a social anthropologist, a social psychologist and two psychoanalysts into the internal and external relationships of twenty London families, with particular reference to the social roles of husband and wife. The families were chosen because they were 'ordinary' in the sense that the research workers went to find them; they did not come to the research team for help, although two of the families had approached another agency for advice. A large element of self-selection necessarily entered into the choice since the inquiry involved the voluntary collaboration of the subjects over a prolonged period. Information was obtained in the course of a dozen or so home interviews lasting an hour or more. This was supplemented, in fifteen of the twenty cases, by clinical interviews with a psychoanalyst at the Tavistock Institute of Human Relations. The present volume is concerned primarily with the social data obtained. It deals with five main topics: (i) The degree of conjugal role segregation shown by the husband and wife, i.e. the extent to which domestic and leisure activities were shared. (ii) The extent to which relationships outside the conjugal family were with a closely knit network of friends, neighbours, and workmates, or were with scattered and unconnected acquaintances. (iii) The degree of intimacy of contact with members of the extended kinship group. (iv) The views of the informants on what constitutes 'normal' behaviour in various aspects of family life. And (v) The views of the informants on questions of social class and status mobility.

The majority of informants had a relatively loose-knit network of friends and acquaintances and little contact with kin. They had difficulty in making articulate their views on social class and the norms of family life. Bott suggests that whereas, in a close-knit network, external sanctions determine attitudes and behaviour directly, in a loose-knit network people do not acquire their values solely by internalizing them from outside, but tend to project their own standards on to others. The principal conclusion emerging from the inquiry is that the more closely connected the family's social network the greater the tendency for husband and wife to have a strict division of labour in the home and to lead more or less separate lives as far as their leisure activities are concerned.

A later book by Dr A. T. M. Wilson will deal with the 'unconscious aspects of the relationship between husband and wife, of the interplay of their personalities, and of the emotional tasks of marriage'. Bott admits that the small size of the sample makes firm conclusions impossible and her findings are put forward as hypotheses which must be tested in the light of further comparative evidence. It is to be hoped that Dr Wilson's forthcoming book will justify the use of intensive interviewing techniques, because none of the topics dealt with in the present volume was outside the scope of an ordinary random-sample survey. In some respects the investigators would have done better if they had relied upon shorter interviews with a larger number of people. What they would have lost in depth would have been amply repaid by having a representative sample.

ANTHONY H. RICHMOND

The Nature of Conflict: Studies on the Sociological Aspects of International Tensions. By The International Sociological Association, in collaboration with Jessie Bernard, T. H. Pear, Raymond Aron and Robert C. Angell. Paris: UNESCO, 1957. Pp. 314. 21*s*.

The statement is made in the Preface to this book that it presents an up-to-date survey and evaluation of research by sociologists and social psychologists into the nature, conditions and implications of human conflict, and particularly conflict between nations. The Introduction is a review of the UNESCO Tensions Project, worked out by the Central Conference, Mexico City, 1947, and of progress made since 1948.

From the point of view of psychology the interest of the second chapter, by Emeritus Professor T. H. Pear, is considerable. He has made a survey and evaluation of publications and researches which seemed to him of special importance, dealing with psychological studies of the tensions leading to conflicts. It is a fair and interesting survey and will be useful to social psychologists.

Prof. Jessie Bernard contributes a chapter on the sociological study of conflict, Prof. Raymond Aron contributes one upon conflict and war from the point of view of historical sociology, and Prof. Robert C. Angell a chapter on the discovery of paths to peace. The last writer takes the view that states which are in interaction will enjoy peace, not so much by the mere absence of war, but only if they are parts of a social system which embraces them.

The book concludes with a useful list of 1160 titles of relevant books and other publications, prepared by the 'Centre d'Études Sociologiques', and a list of periodicals. If the reader feels disappointed because he does not find in the book decisive answers to his biggest questions, it must be realized that such answers cannot yet be given, but he will find a study of many problems and ideas which will help him.

R. W. PICKFORD

Perceptanalysis. By Z. A. Piotrowski. New York and London: The Macmillan Company. 1957. Pp. xvii+505. 47*s*.

Despite the wider implications of its title, Piotrowski's *Perceptanalysis* is concerned entirely with the Rorschach method. As such it invites comparison with Klopfer's *Developments in the Rorschach Technique* (reviewed in this *Journal*, 1955, **46**, 72–3 and 1957, **48**, 150–1).

Lacking experience of testing Piotrowski's rationale, one has to rely on impression which suggests that, while on occasion he shows quite brilliant clinical insight, the work is far below that of the Klopfer group as regards scientific approach. There are numerous cases of *non sequitur*, circular reasoning, and the like, and—as is unfortunately all too common in Rorschach literature—statements are often made quite dogmatically without supporting evidence. One of the few places where detailed figures are quoted is in a discussion of the ranking of plates in terms of initial reaction time; here the issue is confused by the figures being mean ranks whereas the discussion refers to mean reaction times. The interpretation seems to be at variance with the figures, but there is no way of telling whether this is indeed the case.

It is difficult to get an all-over picture of Piotrowski's scoring system, since he gives no summary list of symbols, ratios calculated, etc. Two important innovations, however, stand out. First, his admission of an *MC* response, i.e. one that contributes equally to the human movement and the colour scores. and an extension of this principle to other combinations of two or even three determinant categories. This allows him to abolish the necessity to recognize 'precedence' of one determinant over another, and the often meaningless distinction between 'main' and 'additional' scores of the Klopfer method. Secondly, his treatment of shading responses is entirely original, in that he uses two basic symbols only: *c* for

ight grey and variegated shading' and *c′* for 'dark-shading'. Each of these is also made the subject of weighted score analogous to 'colour sum'. While this treatment obviates many scoring problems, it ails to recognize the interpretative significance of the texture response which many Rorschach workers nd particularly illuminating; the author also weakens his case by the remark (p. 257) that 'Percept-nalysis is not so precise a measuring instrument for a simplification of this degree to cause significant rrors'. In this connexion, too, Piotrowski produces the following 'gift' to those who hold Rorschach jargon' up to ridicule: 'As the pig is a very active animal it also symbolizes the *c′R* tendency while he sheep, which is much less active, symbolizes the *cR*. Liddell reported that Pavlov did not think hat one could develop an experimental neurosis in pigs...(but)...succeeded where Pavlov had failed nd found that pigs become increasingly aggressive and violently active as they are made more and nore uncomfortable by the conditioned reflex experiments. Thus, they behaved in a typical *c′* manner. heep on the other hand became less and less active, more stiff, and developed cardiovascular symptoms s the experimental neurosis grew stronger. Thus, their behaviour was of a *c* type. In order to alleviate nxiety, the pig increased its activity, while the sheep decreased its activity. A human being, having a nore complex central nervous system than either pig or sheep, is capable of both types of reaction.'

A comparable innovation, on the interpretative level, is the recognition of specific 'shock' elements ssociated with each plate. In the reviewer's experience, plate IX frequently produces disturbances in esponse which appear to be due to difficulties in form interpretation rather than to the colour content. To describe such a disturbance simply as 'Plate IX shock' (for which, of course, the author has a rationale) would appear to be an advance on the assumption—which one sometimes suspects has been made—hat because stimulus variables are present they are necessarily operative.

Notwithstanding some of the foregoing criticisms, a fair summing-up would be that this is a stimulating book, with many original and controversial features to assure its welcome. Particularly interesting is he information regarding the background and early history of the technique, including frequent references ack to Rorschach's own writings. Only the out-and-out sceptic will deny that it throws fresh light on he many problems of personality dynamics.

B. SEMEONOFF

Insight: A Study of Human Understanding. By B. J. F. Lonergan. London: Longmans, Green. 1957. Pp. xxx+785. 63*s*.

A long and closely argued philosophical book by the Professor of Dogmatic Theology in a Catholic University is not meat for many psychologists. And in Father Lonergan's book there are lengthy meta-physical stretches which probably no psychologist *qua* psychologist could be troubled to read. Never-heless I found, and I think others will find, much in it to reward them. My admiration increased as I ead, and my final impression was that this is a great book, and a book of which psychologists ought to now.

It is about Insight, the activity of the organizing intelligence. It examines the nature of Insight, aking examples in the first instance from mathematics and physics, with which subjects the author hows himself to have a considerable mastery; it analyses the insights of common sense in a most useful ouple of chapters; it complements the concept of Insight with the concept of Oversight, and discusses he biases of the understanding and the implications of the flight from understanding. It goes on to liscuss the philosophical implications of understanding. The author himself summarizes his theme as ollows: 'Thoroughly to understand what it is to understand, and not only will you understand the broad ines of all that is to be understood, but you will possess a fixed base, an invariant pattern, opening pon all further developments of understanding.'

A book with such a theme may well be regarded as just another instance of the delusion of the omni-otence of thought, and dismissed forthwith by the scientifically minded psychologist. In this case I hink he would have made a mistake. This book shows a sustained level of first-hand philosophical thinking f a high order. It combines contemporary scientific knowledge and an awareness of contemporary roblems with a broad grasp of philosophical tradition. It is cogently reasoned and lucidly written. t is often penetrating and humane in its observations. It can be recommended to those psychologists nterested in cognitive processes who have time and patience, a tolerance for philosophy, and a spare hree guineas—for this is a book which one should possess for oneself if one is to trouble with it at all.

L. S. HEARNSHAW

The Collective Dream in Art. By WALTER ABELL. Harvard: University Press. London: Oxford University Press. 1957. Pp. xv+378+39 Plates. 60*s*.

This is a valuable book deserving serious attention from anyone interested in the function of art and of artists in society and willing to reflect on the possibility of integrating historical, sociological and psychological concepts with the findings of art history in order to understand that function. The author's view, succinctly stated on the inside book cover, is that: 'Arts, myths and other forms of cultural expression are to the society that produces them what dreams are to an individual, i.e. symbolical projections of underlying mental states caused, in turn, by the outward circumstances of life.' Walter Abell, who died before his book appeared, was Professor of Art at Michigan State University. He wrote as a specialist in the history of the visual arts, seeking both to enrich his own field by borrowing from allied disciplines and also to suggest an integrated viewpoint about which he clearly felt deeply.

The author's own discipline has contributed to two aspects of the book distinguishing it pleasantly from the crude psychoanalytic ruminations that psychologists are sometimes asked to accept from artistic and literary colleagues. First, Prof. Abell relates his theory to a specific problem in the history of art, namely that of the grotesque in Gothic sculpture. To this he devotes more than a third of the volume in the spirit of trying out a hypothesis to see if it fits the facts. Secondly, Prof. Abell has been chary of accepting either psychological or historical explanations that claim to be all-inclusive. Though, for example, his psychological borrowings are mainly from Freud and Jung, he steers clear of the biological and cultural narrowness that colours some extreme forms of psychoanalytic orthodoxy. Similarly, when writing of the 'collective dream', Professor Abell does not seem to have lost himself in the metaphysical railway cloakroom of the Jungian collective unconscious nor taken over the curious archetypal luggage supposedly deposited there. Again, while borrowing from history and sociology the view that we should seek in economic and social conditions a partial explanation of the negative, of the positive and of the neutral or accepting attitudes he believes to be externalized in art, Prof. Abell is well aware of the fallacies in Marxism. His approach therefore is critical and in these respects balanced. Whether or not Prof. Abell establishes a case remains a matter for discussion. For example, is his viewpoint really new? Has he, as he himself wonders, relied too heavily on direct projection and repression and allowed compensatory mechanisms too little weight? Is his attempt to balance socio-economic and personal factors in the production of art forms adequate or subtle enough? How many artists would accept his discussion of their role, sympathetic as it is? Though some incidental comments are extremely good, the exposition, especially in Part I, strikes the psychological reader as somewhat laboured. Throughout, though less in later chapters, this reviewer sighed periodically for the sparkle with which a Huizinga or a Paul Hazard would have handled comparable material. Occasionally Prof. Abell seems to write with elation about his approach and provokes the reader to be as chary as the author can be of the easily all-inclusive.

But when all this is said, one remains with the impression of a writer, wide in interest, struggling to be objective, unwilling to tolerate cheap negation, and in love with the arts and honesty even if experience is tragic. This is an honourable impression to have left behind and one can but write of it with respect. The book is well produced and includes more than a hundred photographs. These succeed in providing appropriate illustrations, but the presentation would be more effective if in a second edition the publishers could increase the size of the individual photos and arrange them where relevant in the texts. As it is the book is an asset to any good library.

JOAN WYNN REEVES

Instinct in Man: In the Light of Recent Work in Comparative Psychology. By RONALD FLETCHER. London: George Allen and Unwin. 1957. Pp. 348. 40*s*.

This is an exposition and critique of modern views on instinct, with particular reference to the contributions of ethology and psychoanalysis. Dr Fletcher has evidently read widely in both fields and has thought closely about the theoretical issues involved. Unfortunately, he appears to be without personal experience in either discipline and his exposition in consequence lacks that sense of reality which direct exposure to the facts alone can bring. None the less, his work will be of considerable help to students of psychology and others seeking guidance in this confused and controversial field.

O. L. ZANGWILL

tatistics for the Social Sciences. By T. G. CONNOLLY and W. SLUCKIN. London: Cleaver-Hume. Second ed., 1957. Pp. viii+166. 16*s*.

veryone who has been concerned in the teaching of statistical method to the mathematically unsophisicated has probably, at some time or other, thought of writing his own text. Only in this way can one e sure that the student has before him, in black and white, everything that he ought to have—and as ttle as possible to frighten him. Failing this, such a teacher could do worse than fall back on Drs Conolly and Sluckin, who have had a great deal of relevant experience. Their exposition is usually clear, nd all the basic techniques are adequately covered.

It is easy to criticize a book of this sort on the basis of what, beyond the essentials, one would oneself ave included or omitted. In the present instance, one might question the necessity for more than a assing mention of the mean deviation; the retention of the Spearman 'footrule' in addition to *rho*; and ne desirability of burdening the beginner with the labour of computing an approximate tetrachoric orrelation coefficient by solving an equation involving the first two powers of r_t. On the other hand, ne would perhaps have liked to see included a treatment of point biserial as well as biserial correlation, f the contingency coefficient, and possibly even of two-way problems, as well as one-way, in the new hapter on the analysis of variance. However, one cannot have everything.

It is a pity that a few misprints have been overlooked, including at least one in a formula which is uoted once only. Again, a distinction might have been made between tables which are merely incidental o worked examples, and those to which reference must be made, e.g. in applying tests of significance. he latter are well chosen, and present the required information concisely and economically. Finally, ne could wish to have certain examples (e.g. correlation from a scattergram) worked out more fully, hile others of more purely theoretical interest, such as the derivation of values of ordinates of the ormal curve, would perhaps have been better omitted.

B. SEMEONOFF

rgonomics: Human Factors in Work, Machine Control and Equipment Design. Vol. 1, No. 1. Nov. 1957. London: Taylor and Francis. Pp. 100. 25*s*.

he sampling error that might result from attempting to review a journal on the basis of its first number nakes it unwise to do more than welcome a new contemporary and wish it success. But one of the attracive features of this, as a first number, is that it comes into the world with so little fuss. There is a short ditorial defining the field, followed by eight workmanlike papers which are so much at home with one nother as to give the impression that we are reading a journal whose scope and policy have long been stablished. In this number the authors are well distributed between the U.S., the U.K., and the Coninent, and the papers range from concrete practical studies such as 'Human Limitations and Vehicle Design' by Ross A. McFarland, or 'Factors in Fatigue and Stress in the Operation of High-Speed Diesel Passenger Railway Cars with only One Driver Present' by Robert S. Schwab, to an ingenious mathenatical approach to the estimation of skill by J. R. De Jong. Between these extremes there are studies f working conditions, of training methods, and techniques of validation in terms of user preference. ll in all, this is an excellent beginning, and there seems every reason to predict a long and successful uture.

JAMES DREVER

OTHER PUBLICATIONS RECEIVED

More About the Backward Child. By HERTA LOEWY. London: Staples. 1957. Pp. 138 15*s*.

Fear: Contagion and Conquest. By J. C. MOLONEY. New York: Philosophical Library Pp. 140. $3.75.

The Substitute Teacher as Indicator. By M. KOSKENNIEMI. Helsinki: Finnish Academy of Science and Letters. 1957. Pp. 58. 150 F.mk.

A Course in Multivariate Analysis. By M. G. KENDALL. London: Charles Griffin. 1957 Pp. 185. 22s.

Self-conditioning: A New Way to a Successful Life. By HORNELL HART. London: Old bourne Press. 1957. Pp. 198. 15*s*.

The Child Within the Group: An Experiment in Self-government. By M. E. TURNER Stanford University Press. London: Oxford University Press. 1957. Pp. 93. 24*s*

Stammering and its Treatment. By FRANKLIN BROOK. London: Pitman Medical Publishing Co. 1957. Pp. 134. 21*s*.

Stuttering: A Psychodynamic Approach. By DOMINICK A. BARBARA. London: Hutchinson. 1958. Pp. 255. 25*s*.

Fun Comes First for Blind Slow-learners. By MILDRED B. HUFFMAN. Illinois: Thomas Oxford: Blackwell. 1957. Pp. 157. 37*s*. 6*d*.

Rencontre: Contributions à une Psychologie Humaine Dédicés au Professeur F. J. J. Buytendijk. By VARIOUS AUTHORS. Utrecht: Spectrum. 1957. Pp. 520. 25 H.fl.

Pupils' Arithmetical Ability: Differences due to the Teacher. By T. VAHERVUO. Helsinki Finnish Academy of Science and Letters. 1957. Pp. 48. 200 F.mk.

Introductory Psychology: An Approach for Social Workers. By D. R. PRICE-WILLIAMS London: Routledge and Kegan Paul. 1958. Pp. 203. 18*s*.

Perceptual Development: An Investigation within the Framework of Sensory-Tonic Field Theory. By S. WAPNER and H. WERNER. Clark University Press. 1957. Pp. 95 No price given.

A Study of the Major Psychoses in an English County. By M. SHEPHERD. London Chapman and Hall. 1957. Pp. 159. 25*s*.

Leader Attitudes and Group Effectiveness. By FRED E. FIEDLER. University of Illinoi Press. 1958. Pp. 69. $1.75.

Value in Social Theory: A Selection of Essays on Methodology. By GUNNAR MYRDAL Ed. by P. STREETEN. London: Routledge and Kegan Paul. 1958. Pp. xlvi+269 32*s*.

Emotional Conflict. By PETER FLETCHER. London: Pan Books. 1958. Pp. 158. 2*s*. 6*d*

Prescription for Survival. By BROCK CHISHOLM. Columbia University Press. London Oxford University Press. 1957. Pp. 92. 20*s*.

Mental Deficiency. By J. E. W. WALLIN. Brandon: Journal of Clinical Psychology 1956. Pp. 200. $5.00.

The Uses of Argument. By STEPHEN TOULMIN. London: Cambridge University Press 1958. Pp. 264. 22s. 6*d*.